مجله کنترل
ISSN 2008-8345
جلد ۹، شماره ۴، زمستان ۱۳۹۴، صفحه ۲۵-۱

پیشنهاد توابع فعال‌ساز بازه‌ای در شبکۀ عصبی بر پایه توابع شعاعی برای پیش‌بینی سیستم‌های غیر خطی پویا

اله‌یار ظهوری زنگنه[۱]، محمد تشنه‌لب[۲]، مجتبی احمدیه خانه‌سر[۳]

[۱] دانشجوی دکترای مهندسی کامپیوتر - هوش‌مصنوعی، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران،
z.zangeneh@gmail.com

[۲] استاد، قطب علمی کنترل صنعتی، دانشکده مهندسی‌برق، گروه کنترل، دانشگاه صنعتی خواجه نصیرالدین طوسی، teshnehlab@eetd.kntu.ac.ir

[۳] استادیار، دانشکده برق و کامپیوتر، گروه مهندسی قدرت و کنترل، دانشگاه سمنان، ahmadieh@semnan.ac.ir

(تاریخ دریافت مقاله ۱۳۹۴/۷/۲۳، تاریخ پذیرش مقاله ۱۳۹۴/۱۰/۴)

چکیده: «شبکۀ عصبیِ بر پایۀ توابعِ شعاعی»[۱] یک تقریب‌گر عمومی می‌باشد. در این مقاله «تابعِ فعال‌سازِ گرانولی»[۲] برای بهبودِ یادگیری این شبکه در نویزی پیشنهاد می‌گردد که یک تابعِ گاوسی با «انحراف استاندارد بازه‌ای و میانگین ثابت» است و به آن «تابعِ فعال‌سازِ بازه‌ای» نیز گفته می‌شود. در لایۀ میانیِ این شبکه، سه پارامترِ وابسته به توابعِ فعال‌سازِ گرانولی آموزش می‌بینند که «مرکزِ توابعِ فعال‌سازِ گرانولی» که مرکز دسته نامیده می‌شود، کرانِ پائینِ انحرافِ استاندارد و کرانِ بالایِ انحرافِ استاندارد این توابع می‌باشند. در لایۀ خروجی دو پارامتر دیگر یعنی «مرکزِ وزن‌هایِ بازه‌ای» و «بازۀ این وزن‌ها» آموزش می‌بینند. برای آموزش این پارامترها از روش «الگوریتم خوشه‌بندی K-Means» استفاده شده است. در این روش، آموزش شبکه در راستای «گرانوله‌سازیِ پائین به بالا»[۳] می‌باشد که در آن بردارهای ورودی به شکل گرانول‌های بزرگتر در لایۀ میانی خوشه‌بندی می‌گردند. از روش «گرادیان نزولی» نیز برای آموزش پارامترهای شبکه استفاده شده و نتایج با روش جدید مقایسه گردیده است. عملکردِ این شبکه با شناسایی «یک سیستمِ غیر خطیِ پویایِ U شکل با پنج ورودی»[۴] و پیش‌بینیِ «سریِ زمانیِ آشوبِ مکی گلاس»[۵] در شرایط نویزی و بدون نویز سنجیده می‌شود. از نتایج معلوم می‌گردد که استفاده از تابعِ فعال‌سازِ گرانولی در ساختار شبکۀ عصبیِ RBF، باعث کاهش حساسیت به تغییرات ورودی شده و عملکرد آن در شرایط نویزی بهبود می‌یابد.

کلمات کلیدی: شبکۀ عصبیِ بر پایۀ توابعِ شعاعیِ گرانولی[۶]، تابعِ فعال‌سازِ گرانولی، داده‌هایِ نویزی، انحرافِ استانداردِ بازه‌ای، سیستم‌هایِ غیرِ خطیِ پویا و توابعِ آشوب.

Proposing Interval Activation Functions in Radial Basis Function Neural Network to Predict Nonlinear Dynamic Systems

Allahyar Zohoori Zangeneh, Mohammad Teshnehlab, Mojtaba Ahmadieh Khanesar

Abstract: A Radial Basis Function Neural Network (RBFNN) is a general approximator. In this paper a granular activation function is proposed to improve its learning under the noisy conditions. The granular activation function is also named the interval activation function and it is typicaly the Gaussian function which benefits from having a fixed mean and an uncertain standard deviation. The hidden layer of the proposed network has a total of three parameters to train that it consists the means, the lower bounds of the standard deviations and the higher bounds of the standard deviations of the Gaussian functions. The output layer parameters for training are the means of the interval weights and the intervals of the weights. "K-Means clustering algorithm" method is used to train

[1] Radial Basis Function Neural Network (RBFNN)
[2] Granular activation functions
[3] Bottom-up granulation
[4] Nonlinear dynamic system with multiple time delays
[5] Mackey glass chaotic time series
[6] Granular Radial Basis Function Neural Network (GRBFNN)

these parameters. The purpose of the above learning method is regarded as one of the granular method presenting the bottom-up granulation which causes the input vectors clustered in the larger granules in the hidden layer. Gradient descend method is also used to train these parameters to compare with this novel method. The structure is tested with or without noisy data to identify a nonlinear dynamic system with multiple time delays and to predict a chaotic model, Mackey-Glass. It has been shown that the sensibility related to input alterations reduces because of using the granular activation function in RBF Neural Network structure and the response of Granular RBF Neural Network with noisy data is better than RBF Neural Network.

Keywords: Granular Radial Basis Function Neural Network (GRBFNN), Granular activation function, Noisy data, Interval standard deviation, Nonlinear dynamic systems and Chaotic Models.

۱- مقدمه

یک تابع شعاعی[1] تابعی است که مقدار آن فقط به فاصله از یک نقطه که مرکز دسته می‌باشد وابسته است. این توابع در تخمین[2] [۲]، پیش‌بینی سری‌های زمانی[3] [۱، ۴۳] و کنترل[۲۲] مورد استفاده قرار می‌گیرند. در یک شبکهٔ عصبی، از این توابع می‌توان، به عنوان تابع فعال‌سازی نُرون استفاده کرد. یک شبکهٔ عصبی بر پایه توابع شعاعی در حالت کلی شامل سه لایه می‌باشد، لایهٔ ورودی، لایهٔ میانی یا همان لایهٔ پنهان[4] و لایهٔ خروجی[۱، ۲۲، ۲۶].

در این مقاله یک شبکهٔ عصبی RBF گرانولی، برای پردازش سیستم‌های همراه با نویز معرفی می‌گردد. توابع دارای «میانگین بازه‌ای و انحراف استاندارد ثابت[5]» و یا «انحراف استاندارد بازه‌ای و میانگین ثابت[6]»، دو نوع تابع فعال‌ساز گرانولی می‌باشند[۶، ۷، ۲۳]. یک شبکهٔ عصبی RBF گرانولی، توابع فعال‌ساز گاوسی گرانولی با «انحراف استاندارد بازه‌ای و میانگین ثابت» را، در یک سیستم عصبی وارد می‌نماید تا مقاومت در برابر نویز آن را افزایش دهد[۶، ۷، ۱۰، ۱۱، ۱۸، ۲۰].

یک سیستم عصبی گرانولی روشی مناسب‌تر نسبت به دیگر شبکه‌های عصبی در بررسی عدم قطعیت‌ها می‌باشد[۵]. زیرا در صورت وجود نویز در مقادیر ورودی، پایداری سیستم را در مقابل آن افزایش می‌دهد[۶، ۷، ۱۰].

یک شبکهٔ عصبی RBF گرانولی، می‌تواند به عنوان کلاسی از شبکه‌های تطبیق‌پذیر محسوب گردد. شبکه‌های عصبی تطبیق‌پذیر، توسعه یافته شبکه‌های پیش‌رو هستند که تابع فعال‌ساز نُرون‌های آن می‌تواند به هر فرمی باشد یعنی لزومی ندارد به صورت سیگموید، سینوسی یا گاوسی باشد[۲]. این کلاس از شبکه‌های تطبیق‌پذیر در کاربردهای شناسایی، پیش‌بینی و کنترل سیستم‌ها استفاده شده و قابلیت خود را به خوبی نشان داده است. به سبب قابلیت‌های خوب این شبکه‌ها، می‌توان از آن به‌عنوان یک مُدل کنندهٔ خطی سازی محلّی و یک روش خطی سازی متعارف در تخمین حالت و کنترل نیز نام برد[۲].

روش آموزش پارامترها در شبکهٔ عصبی RBF گرانولی یک مسأله با اهمیّت است. یکی از این روش‌ها بر مبنای «گرادیان نزولی[7]» بنا شده است، که محاسبهٔ مشتق در بعضی مراحل آن در قانون زنجیره‌ای[8] بسیار دشوار است. از طرفی به کار بردن آن ممکن است ما را در مینیمم محلی[9] گرفتار سازد[۳۵، ۳۶]. پیچیدگی «الگوریتم گرادیان نزولی» برای آموزش پارامترها در شبکهٔ عصبی RBF و شبکهٔ عصبی RBF گرانولی فارغ از تعداد ورودی‌ها از درجهٔ ۴ یعنی $O(m \times n \times T \times Epoch)$ می‌باشد، که m تعداد نُرون‌های لایهٔ میانی، n ابعاد ورودی‌ها، T تعداد نمونه‌های ورودی و $Epoch$ تعداد دفعات تکرار الگوریتم است[۳۵]. در این پژوهش سعی شده است یک روش دیگر که بتواند پارامترها را سریع‌تر و آسان‌تر از «گرادیان نزولی» آموزش دهد معرفی گردد. زیرا در روش «گرادیان نزولی»، همگرایی پارامترها به شدت وابسته به نرخ آموزشِ[10] مناسب است که اغلب یافتن آن بسیار دشوار می‌باشد[۳۶].

در روش پیشنهادی این مقاله که «الگوریتم خوشه‌بندی K-Means[11]» نامیده می‌شود همگرایی پارامترها سریع‌تر و در تعداد دفعاتِ تکرار(مرحله)[12] کمتری[۲۳، ۳۷، ۴۴] صورت می‌گیرد. پیچیدگیِ «الگوریتم خوشه‌بندی K-Means» برای آموزش پارامترها در شبکهٔ عصبیِ RBF گرانولی بدون توجه به تعداد ورودی‌ها از درجهٔ ۳ یعنی $O(m \times T \times Epoch)$ می‌باشد، که m تعداد خوشه‌ها(یا همان تعداد نُرون‌های لایهٔ میانی)، T تعداد کلّ اشیاء(یا همان تعداد نمونه‌های ورودی) و $Epoch$ تعداد دفعات تکرار الگوریتم است[۳۷]. آموزش پارامترهای لایهٔ میانی بوسیله‌ی روش «الگوریتم خوشه‌بندی K-Means»،

[1] Radial basis
[2] Estimation
[3] Time series prediction
[4] Hidden Layer
[5] Granular activation function having an uncertain mean and fixed standard deviation σ
[6] Granular activation function having an uncertain standard deviation σ and fixed mean

[7] Gradient descend
[8] Chain rule
[9] Local minimum
[10] Learning rate
[11] K-Means clustering algorithm
[12] Epoch

انعطاف‌پذیری سیستم را نسبت به آموزش با روش «گرادیان نزولی» افزایش می‌دهد[۲۳، ۳۷].

از طرفی «الگوریتم خوشه‌بندی K-Means» در شناسایی و پیش‌بینیِ توابع آشوب با محدودیت‌هایی روبرو است؛ زیرا دو بردار ورودی که فاصلهٔ اقلیدسی[۱] کوچکی دارند و ممکن است در لایهٔ پنهان در یک خوشه[۲] قرار گیرند، می‌توانند در این گونه توابع خروجی‌هایِ دور از هم داشته باشند. مگر این که افق پیش‌بینی را در این توابع به قدری کوچک در نظر بگیریم تا دو بردار ورودی که فاصلهٔ اقلیدسی کوچکی دارند، خروجی‌های نزدیک به هم تولید کنند[۲۷ ۲۸]. به همین دلیل، بکار بردن «الگوریتم خوشه‌بندی K-Means» سبب می‌گردد که وزن‌های لایهٔ خروجی بار بیشتری را در شناسایی و پیش‌بینیِ توابع آشوب تحمل نمایند[۳۷]. ولی در مورد شناساییِ سیستم‌های غیر خطیِ پویا که آشوبی نیستند، «الگوریتم خوشه‌بندی K-Means» در تعداد تکرار کمتر عملکرد بهتری دارد.

عملکرد شبکهٔ عصبیِ RBF گرانولی، بوسیله شناساییِ «یک سیستمِ غیر خطیِ پویا با پنج ورودی[۶]» و پیش‌بینیِ «سری زمانیِ آشوبِ مکی گلاس[۳] [۳۰]» مورد آزمایش قرار می‌گیرد[۲۷]. نویز به کار رفته دارای «نسبت سیگنال به نویز[۴]» برابر صفر، پنج و ده خواهد بود که با شرایط بدون نویز مقایسه می‌گردد[۱۱].

از آنجا که در رابطه با آموزش پارامترهای لایهٔ میانیِ شبکهٔ عصبیِ RBF گرانولی، به کمک «الگوریتم خوشه‌بندی K-Means» و «گرادیان نزولی» در شرایط نویزی و بدون نویز و مقایسهٔ آنها با یکدیگر و با دیگر شبکه‌هایِ عصبی و عصبی فازی، پژوهشی انجام نشده است، این مقاله به‌منظور پاسخ به پرسش‌های زیر شکل گرفته است:

- شبکهٔ عصبیِ RBF گرانولی، با آموزش پارامترهای لایهٔ میانی بوسیلهٔ دو روش «الگوریتم خوشه‌بندی K-Means» یا «گرادیان نزولی»، و پارامترهای لایهٔ خروجی بوسیله‌ی روش «گرادیان نزولی» نسبت به شبکهٔ عصبیِ RBF با آموزش همهٔ پارامترها بوسیلهٔ روش «گرادیان نزولی»، در کدام یک از شرایط نویزی و بدون نویز بهتر عمل می‌کند؟
- ایجاد یک ساختار بازه‌ای در دو پارامتر الف) انحرافِ استانداردِ توابع فعال‌ساز گرانولی که یک پارامتر غیرخطی لایهٔ میانی است و ب) وزن‌های لایهٔ خروجی که یک پارامتر خطی لایهٔ خروجی است، چه تأثیری در افزایش کارایی سیستم در شرایط نویزی خواهد داشت؟
- در شبکهٔ عصبیِ RBF گرانولی، آموزش پارامترهای غیرخطی لایهٔ میانی بوسیله‌ی روش «الگوریتم خوشه‌بندی K-Means» در مقایسه با «الگوریتم گرادیان نزولی» چه تفاوتی در شناساییِ سیستم‌هایِ غیر خطیِ پویا و پیش‌بینیِ توابعِ آشوب دارد؟ این تفاوت در خطایِ شناسایی و پیش‌بینی و زمان اجرایِ برنامه چگونه خواهد بود؟
- کاراییِ شبکهٔ عصبیِ RBF گرانولی در مدیریت عدم‌قطعیتِ ناشی از شرایط نویزی؛ نسبت به دیگر شبکه‌های عصبی و عصبی فازی چگونه است؟ آیا به کار بردن شبکهٔ عصبیِ RBF گرانولی نسبت به دیگر سیستم‌های هوشمند دارای مزیت است؟

۲- مرور کارهایِ انجام شده در زمینه نویز و کاهش اثرآن[۵]

۲-۱ تعریف نویز

در یک تعریف کلی، به هر نوسان و تغییر ناخواسته که بر روی سیگنال‌های مورد اندازه‌گیری ظاهر شود، نویز گفته می‌شود. هر کمّیتی می‌تواند نویزی گردد[۱۴، ۱۶].

- در مدارهای الکتریکی بیشتر با نویز ولتاژ و جریان سر و کار داریم؛ این نویز ناشی از تغییراتِ دماییِ محیطِ انتقال انرژی و تاثیر آن بر روی حرکت الکترون‌ها است.
- در حوزهٔ امواج رادیویی و مایکروویو[۶] با نویزهای الکترومغناطیسی و گاهی نیز با نویزی که ناشی از گرما یا تابش[۷] و یون‌های کم انرژی باشد روبرو هستیم.

در هر آزمایش دقیق و با کیفیت بالایی که انجام می‌شود، باید بتوان نویز محیط را پیش‌بینی و تأثیر آن را کم کرد. اهمیّت تحلیل نویز هنگامی کاملا نمایان می‌شود که متوجه می‌شویم کیفیت سیگنال اندازه‌گیری شده تنها به مقدار انرژی سیگنال بستگی ندارد بلکه به وسیلهٔ «نسبت سیگنال به نویز» تعیین می‌شود . نتیجه تحقیقات نشان می‌دهد که بهترین روش برای بهبود «نسبت سیگنال به نویز»، کاهش نویز است نه افزایش قدرت سیگنال[۱۰، ۱۱، ۱۶].

۲-۲ فرآیند تولید نویز

نویز طبق تعریف، غیر قابل کنترل است و مقدار دقیق آن در آزمایش‌های گوناگون با هم متفاوت است. پس در واقع فرآیند تولید نویز از نوع فرآیندهای تصادفی است و معمولاً تابع توزیع احتمال[۸] [۳۹] متغیّر تصادفی نویز را بر اساس قضیهٔ حد مرکزی[۹] [۳۹] به شکل گاوسی با

[۱] Euclidean distance
[۲] Cluster
[۳] Mackey glass chaotic time series
[۴] Signal to Noise Ratio(SNR)

[۵] Noise Reduction
[۶] Microwave
[۷] Radiation
[۸] Probability Distribution Function (PDF)
[۹] نظریه حدّ مرکزی بیان می‌کند که اگر تعداد نمونه‌های یک متغیر تصادفی به سمت بی‌نهایت میل کند، تابع توزیع احتمال آن متغیر تصادفی به سوی توزیع نُرمال میل می‌کند.

نرمال در نظر می‌گیرند. البته در شرایطی که تعداد نمونه‌های «متغیرِ تصادفی» نویز کم باشد، ممکن است توزیع‌های دیگری نیز مدّ نظر قرار گیرد.

تصادفی بودن نویز سبب می‌شود که تابع توزیع احتمال آن از نوع «نُرمال یا گاوسی با میانگینِ صفر» و به صورت $\mathcal{N}(0,\sigma_n^2)$ در نظر گرفته شود[۹]. بنابراین برای توصیف نویز از مقادیر مربع آن استفاده می‌شود. مقدار مؤثر[1] نویز از جذر میانگین مربعات[2] آن به دست می‌آید. البته این پارامتر هیچ اطلاعاتی در مورد چگونگی تغییر مقدار نویز با زمان و یا اجزای فرکانسی آن نمی‌دهد. اگر ویژگی‌های آماری نویز مانند واریانس[3] یا انحرافِ استاندارد و یا مقدار مؤثر آن با زمان تغییر نکند به آن نویز ایستا[4] گفته می‌شود[۹].

در سیستم هایی که چند منبع نویز وجود داشته باشد نویز کلی می‌تواند به صورت مجموع نویزهای مختلف در نظر گرفته شود. اگر این نویزها مستقل[5] از یکدیگر باشند می‌توان مقدار مؤثر را به صورت جمع مقدارهای مؤثر تک تک منابع نویز در نظر گرفت[۳۹ ۴۱].

۲-۳ انواع نویز

نویزها بیشتر بر اساس تغییرات زمانی و فرکانسی خود از یکدیگر متمایز می‌شوند. در شبکهٔ عصبی بر پایه توابعِ شعاعیِ گرانولی، با نویزهایی سروکار داریم که از نوع سیگنال‌های «گسستهٔ در زمان تصادفی»[6] می‌باشند. یک گروه مهم از این سیگنال‌ها، سیگنال‌های WSS[7] است[۴۳]. یک سیگنال «گسستهٔ در زمان تصادفی»، مانند دنبالهٔ نویزی $\left(n_r^{(t)}\right)_{r\epsilon N}$ را یک سیگنال WSS گویند اگر داشته باشیم:

$$E\left(n_r^{(t)}\right) = a\ constant\ value \quad \forall r\epsilon[-\infty \quad \infty] \qquad (۱)$$

$$E\left(n_p^{(t)} n_q^{(t)}\right) = E\left(n_{p+r}^{(t)} n_{q+r}^{(t)}\right), \quad \forall p,q,r\epsilon N \qquad (۲)$$

از فرمول‌های (۱) و (۲) نتیجه می‌گیریم که برای WSS بودن یک سیگنال «گسستهٔ در زمان تصادفی»، مانند دنبالهٔ نویزی $\left(n_r^{(t)}\right)_{r\epsilon N}$، میانگین آن به ازای هر مقدار r باید ثابت باشد.

$$E\left(\left(n_r^{(t)}\right)^2\right) = a\ constant\ value \Rightarrow Var\left(n_r^{(t)}\right) = a\ constant\ value \qquad (۳)$$

از فرمول (۳) نتیجه می‌گیریم که برای WSS بودن یک سیگنال «گسستهٔ در زمان تصادفی»، مانند دنبالهٔ نویزی $\left(n_r^{(t)}\right)_{r\epsilon N}$، واریانس آن به ازای هر مقدار r باید ثابت باشد. برخی از عمومی‌ترین نویزهای موجود، عبارتند از:

- **نویز سفید**[8]. به یک دنبالهٔ نویزی مانند $\left(n_r^{(t)}\right)_{r\epsilon N}$، نویز سفید گفته می‌شود اگر متغیر تصادفی $n_r^{(t)}$ در این دنباله، دارای میانگین صفر بوده و داشته باشیم:

$$E\left(\left(n_r^{(t)}\right)^2\right) = Var\left(n_r^{(t)}\right) = \sigma_n{}^2 = a\ constant\ value \qquad (۴)$$

تابع چگالی توان[9] و یا طیف توان[10] نویز سفید به فرکانس آن بستگی ندارد و دارای دامنهٔ ثابتی برابر $\sigma_n{}^2$ است که به آن توان نویز گفته می‌شود. البته این یک تعریف ایده‌آل است زیرا اگر از یک عدد ثابت نسبت به فرکانس انتگرال بگیریم، واریانس نویز (یا همان توان نویز) بی‌نهایت به دست می‌آید. نویز سفید به دو صورت ظاهر می شود؛ نویز دمایی[11] و اثر ساچمه‌ای[12].

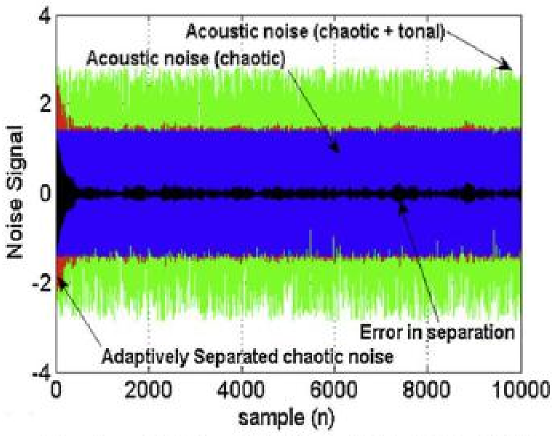
- **آشفتگیِ هارمونیک**[13]. آشفتگی‌های هارمونیک در واقع نویزهای تصادفی نیستند بلکه آشفتگی‌هایی هستند که از منابع نزدیک بر روی سیستم افتاده است. این نویزها می‌توانند به وسیله طراحی‌های مناسب حذف شوند. روش‌هایی که برای حذف این نویز استفاده می‌شوند عبارتند از؛ پوشش محافظ[14]، زمین کردن[15] مناسب و کاهش حساسیت سیستم[16] به نویز. از آن‌جا که آشفتگی‌های هارمونیک دارای فرکانس‌های مشخصی هستند، باعث ایجاد نوسانات نامیرا[17] در سیگنال و ایجاد ضربه[18] در طیف فرکانسی[19] می‌شوند. این رفتار تکین[20] باعث می‌شود که نوع آن‌ها با نویزهای دیگر فرق کند.

- **نویز صورتی[21] یا نویز $\frac{1}{f}$**. در بررسی سیستم‌ها، نویز واقعی سفید نیست بلکه «صورتی» است. به این معنا که دارای فرکانس قطع است. این فرکانس قطع باعث می‌شود که واریانس نویز محدود شود. طیف توان این نویز با آهنگِ $\frac{1}{f}$ کاهش پیدا می کند. توان نویز $\frac{1}{f}$ بستگی به نحوه تولید آن دارد و از وسیله‌ای به وسیله دیگر متفاوت است.

- **نویز آشوبی**[22]. این نویز می‌تواند توسط ماشین‌هایی که دارای قسمت گردنده[23] و یا لبه‌های تیز[24] می‌باشند تولید گردد. معمولاً این نویز

[1] Effective
[2] Root Mean Square (RMS)
[3] واریانس معرف انرژی نویز است
[4] Static
[5] نویزهایی مستقل هستند که میانگین حاصل‌ضرب دو به دوی نویزها صفر شود
[6] Stochastic Discrete Time Signal
[7] Wide-Sense Stationary
[8] White Noise
[9] Power Spectral Density
[10] Power Spectrum
[11] Thermal Noise
[12] Shot Noise
[13] Harmonic Disturbance or Harmonic Oscillation
[14] Shield
[15] Earthing
[16] System Sensitivity Reduction
[17] Undamped Oscillations
[18] Impulse
[19] Frequency Spectrum
[20] Singularity Behavior
[21] Pink Noise or Flicker Noise
[22] Chaotic Noise
[23] Rotating part
[24] Blades

با یک تُن صدا[1] ترکیب می‌گردد که وابسته به سرعت چرخش قسمت گردان ماشین می‌باشد. بخش آشوبی نویز مربوط به برخورد لبه‌های تیز به هوای اطراف می‌باشد[۹].



شکل ۱. نمودار تغییرات نویز حاصل از ترکیب بخش آشوبی و تُن صدا، پیش و پس از جداسازی این دو بخش، در دامنهٔ زمان[۹]

۴-۲ کاهش اثر نویز

پدیده نویز در کنترل سیستم‌های مختلف مشکلی عمومی است و سیستم‌های هوشمند بویژه سیستم‌های هوشمندِ نوع۲[2] [۶، ۷، ۱۱] که در آن‌ها مدیریت عدم قطعیت به خوبی صورت می‌پذیرد، در این حیطه وارد شده‌اند. تشخیص نویز[3] و حذف نویز[4] در بسیاری از زمینه‌ها مانند پردازش تصویر در کنار کاهش اثر نویز از اهمیّت بالایی برخوردارند. با وجود پیشرفت در طراحی شبکه‌های عصبی، ارائه روش‌هایی که نیاز به دورهٔ آموزش کمتری دارند هنوز مفید است. برخی از این روش‌ها در این‌جا آورده شده است:

- استفاده از فیلتر غیرخطی «کاهش نویزِ ضربهٔ فازی[5]» [۱۳، ۱۴].
- به کار بردن روش موجک[6] [۱۲].
- الگوریتم فیلتر خطی FXLMS[7] [۷].
- الگوریتم کنترل با فیلتر غیرخطی VFXLMS[8] [۱۵].
- انواع شبکه‌های عصبی مانند شبکهٔ عصبیِ RBF گرانولی [۱۸، ۲۰] و شبکهٔ عصبیِ ارتباط تابعی[9] [۹].
- سیستم‌های منطق فازی نوع۲[10] [۱۰، ۱۱] که به عنوان نمونه می‌توان با شبکهٔ عصبی فازی نوع۲ بازه‌ای[11] [۶، ۷] آن را پیاده‌سازی نمود.

[1] Tonal
[2] Type 2 Intelligenet Systems
[3] Noise Detection
[4] Noise Canceling
[5] Fuzzy Impulse Noise Reduction Method (FINRM)
[6] Wavelet
[7] Filtered-X Least Mean Square
[8] Volterra Filtered-X Least Mean Square
[9] Functiona Link Artificial Neural Network (FLANN)
[10] Type-2 Fuzzy Logic Systems (T2FLSs)
[11] Interval Type-2 Fuzzy Neural Network

۳- معرفی شبکهٔ عصبیِ RBF

ماتریس n سطری $X^{(t)}$ ورودی شبکهٔ عصبی است که به شکل یک دسته دادهٔ ورودی به آن وارد می‌شود. و ماتریس $\mathcal{N}^{(t)}$، نویز اضافه شده به این دسته دادهٔ ورودی می‌باشد. ماتریس $w^{j,(t)}$ شامل وزن‌های لایهٔ پنهان شبکه و ماتریس‌های $c^{j,(t)}$ و $\sigma^{j,(t)}$ به ترتیب شامل مقادیر مرکز و انحراف استانداردِ «توابعِ فعال‌سازِ گاوسیِ نُرون زاُم لایهٔ پنهانِ شبکه[12]»، می‌باشند. اندیس‌های j، r و t به صورت زیر تعریف می‌گردند[۲۳]:

- برای تعداد خوشه‌ها[13] یا تعداد نُرون‌های لایهٔ میانی داریم: $\mathbf{j = 1,2, \dots, m}$
- برای ابعاد ورودی‌ها داریم: $\mathbf{r = 1,2, \dots, n}$
- برای شماره نمونه‌های ورودی داریم: $\mathbf{t = 1,2, \dots, T}$

هم‌چنین مقادیر اولیه‌ی مرکزِ توابعِ فعال‌سازِ گاوسیِ نُرون‌هایِ لایهٔ پنهان، یعنی $c^{j,(0)}$ معادل $w^{j,(0)}$ در نظر گرفته می‌شود. تابع ψ تابعِ فعال سازِ گاوسیِ نُرون‌هایِ لایهٔ میانی است. نُرم بکار رفته در این‌جا نُرم اقلیدسی می‌باشد. خروجی اولیه‌ی نُرون زاُمِ لایهٔ میانی یعنی $o^{j,(0)}$ را می‌توان به وسیله فرمول زیر به‌دست آورد[۲۳، ۲۶]:

$$o^{j,(0)} = \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(0)}+n_r^{(0)}-c_r^{j,(0)}}{\sigma_r^{j,(0)}}\right)^2\right] = \psi\left[\sum_{r=1}^{n}\left(u_r^{j,(0)}\right)^2\right] = \psi\left(d^{j,(0)}\right) = exp\left(-\frac{1}{2}d^{j,(0)}\right) \quad (۵)$$

شکل ماتریسی متغیرهای به کار رفته عبارت است از:

$$X^{(t)} = \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \\ \vdots \\ x_n^{(t)} \end{bmatrix}_{n\times 1}, \quad \mathcal{N}^{(t)} = \begin{bmatrix} n_1^{(t)} \\ n_2^{(t)} \\ \vdots \\ n_n^{(t)} \end{bmatrix}_{n\times 1}, \quad c^{j,(t)} = \begin{bmatrix} c_1^{j,(t)} \\ c_2^{j,(t)} \\ \vdots \\ c_n^{j,(t)} \end{bmatrix}_{n\times 1} = w^{j,(t)} = \begin{bmatrix} w_1^{j,(t)} \\ w_2^{j,(t)} \\ \vdots \\ w_n^{j,(t)} \end{bmatrix}_{n\times 1}, \quad \sigma^{j,(t)} = \begin{bmatrix} \sigma_1^{j,(t)} \\ \sigma_2^{j,(t)} \\ \vdots \\ \sigma_n^{j,(t)} \end{bmatrix}_{n\times 1} \quad (۶)$$

برای مقادیر وزن‌های اتصال بین لایهٔ ورودی و لایهٔ پنهان شبکه، ماتریس زیر را خواهیم داشت:

$$W^{(t)} = [w^{1,(t)} \quad w^{2,(t)} \quad \dots \quad w^{m,(t)}]_{1\times m} = \begin{bmatrix} w_1^{1,(t)} & w_1^{2,(t)} & \dots & w_1^{m,(t)} \\ w_2^{1,(t)} & w_2^{2,(t)} & \dots & w_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ w_n^{1,(t)} & w_n^{2,(t)} & \dots & w_n^{m,(t)} \end{bmatrix}_{n\times m} \quad (۷)$$

برای مقادیر مرکزِ توابع فعال‌سازِ نُرون‌های لایهٔ پنهان، ماتریس زیر را خواهیم داشت:

[12] Gaussian activation functions of the j[th] neuron in the hidden layer of the neural network
[13] Clusters

$$C^{(t)} = [c^{1,(t)} \quad c^{2,(t)} \quad \dots \quad c^{m,(t)}]_{1\times m} = \begin{bmatrix} c_1^{1,(t)} & c_1^{2,(t)} & \dots & c_1^{m,(t)} \\ c_2^{1,(t)} & c_2^{2,(t)} & \dots & c_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ c_n^{1,(t)} & c_n^{2,(t)} & \dots & c_n^{m,(t)} \end{bmatrix}_{n\times m} \quad (۸)$$

و برای مقادیر انحرافِ استاندارد این نُرون‌ها، ماتریس زیر را خواهیم داشت:

$$\Sigma^{(t)} = [\sigma^{1,(t)} \quad \sigma^{2,(t)} \quad \dots \quad \sigma^{m,(t)}]_{1\times m} = \begin{bmatrix} \sigma_1^{1,(t)} & \sigma_1^{2,(t)} & \dots & \sigma_1^{m,(t)} \\ \sigma_2^{1,(t)} & \sigma_2^{2,(t)} & \dots & \sigma_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^{1,(t)} & \sigma_n^{2,(t)} & \dots & \sigma_n^{m,(t)} \end{bmatrix}_{n\times m} \quad (۹)$$

خروجی لایهٔ پنهان شبکهٔ عصبی برای یک دسته دادهٔ ورودی برابر خواهد بود با:

$$O^{(t)} = [o^{1,(t)} \; o^{2,(t)} \dots o^{m,(t)}] = [\psi(X^{(t)}, \mathcal{N}^{(t)}, c^{1,(t)}, \sigma^{1,(t)}) \;\; \psi(X^{(t)}, \mathcal{N}^{(t)}, c^{2,(t)}, \sigma^{2,(t)}) \dots \dots \psi(X^{(t)}, \mathcal{N}^{(t)}, c^{m,(t)}, \sigma^{m,(t)})] \quad (۱۰)$$

$$O^{(t)} = \left[\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{1,(t)}}{\sigma_r^{1,(t)}}\right)^2\right] \;\; \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{2,(t)}}{\sigma_r^{2,(t)}}\right)^2\right] \dots \dots \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{m,(t)}}{\sigma_r^{m,(t)}}\right)^2\right]\right] \quad (۱۱)$$

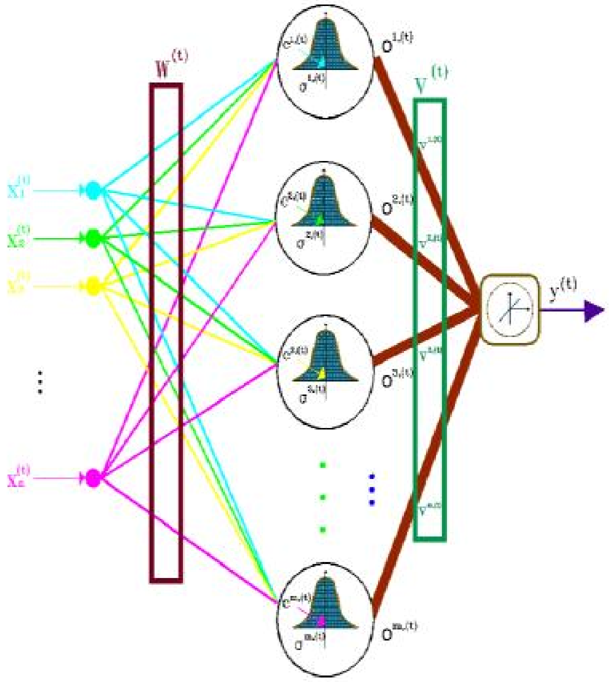
اکنون خروجی نهایی شبکهٔ عصبی برای یک نمونهٔ دادهٔ ورودی برابر خواهد بود با[۲۳]:

$$y_{\mathcal{N}}^{(t)} = \sum_{j=1}^{m}\left(v^{j,(t)} \times o^{j,(t)}\right) \quad (۱۲)$$

آنجا که $v^{j,(t)}$، وزن اتصال بین لایهٔ میانی و لایهٔ خروجی برای j اُمین نُرون لایهٔ پنهان است و یک پارامتر خطی می‌باشد. در نهایت در یک فرمول کلی برای هر یک از $t = 1,2,\dots,T$ نمونهٔ دادهٔ ورودی خواهیم داشت[۲۳]:

$$y_{\mathcal{N}}^{(t)} = G\left(X^{(t)}, \mathcal{N}^{(t)}, C^{(t)}, \Sigma^{(t)}, V^{(t)}\right) = \sum_{j=1}^{m}\left[v^{j,(t)} \times \psi\left(X^{(t)}, \mathcal{N}^{(t)}, c^{j,(t)}, \sigma^{j,(t)}\right)\right] = \sum_{j=1}^{m}\left(v^{j,(t)} \times o^{j,(t)}\right) \quad (۱۳)$$

که $V^{(t)} = \begin{bmatrix} v^{1,(t)} \\ v^{2,(t)} \\ \vdots \\ v^{m,(t)} \end{bmatrix}_{m\times 1}$ ماتریس وزن‌های اتصال بین لایهٔ میانی و لایهٔ خروجی می‌باشد[۲۳].



شکل ۲. یک شبکهٔ عصبی RBF

۳-۱ روش‌های آموزش بکار رفته در شبکهٔ عصبی RBF

برای آموزش پارامترهای غیرخطیِ لایهٔ میانی که «مرکز» و «انحراف استاندارد» توابع فعال‌ساز گاوسی می‌باشند می‌توان از دو روش آموزش یکی از نوع آموزش بی‌سرپرست[1] [۳۳] یعنی «الگوریتم خوشه‌بندی K-Means با m خوشه»[۲۳، ۳۷] و دیگری از نوع آموزش با سرپرست[2] [۳۳] یعنی «گرادیان نزولی» استفاده نمود[۲۳، ۳۵].

۳-۲ الگوریتم خوشه‌بندی K-Means با m خوشه

گام‌های الگوریتم برای آموزش «مرکز توابع فعال‌ساز گاوسی» به شرح زیر است و کاملاً مشابه الگوریتم آموزش «انحراف استاندارد آن‌ها» می‌باشد، جز آنکه در الگوریتم آموزش «انحراف استاندارد»، به جای «مرکز توابع فعال‌ساز گاوسی، یعنی $c_r^{j,(t)}$»، «انحراف استاندارد» یعنی $\sigma_r^{j,(t)}$ جایگزین می‌گردد[۲۳، ۳۷]:

۱. به دو پارامتر «مرکز» یا «انحراف استاندارد» توابع فعال‌ساز گاوسی مربوط به نُرون‌های لایهٔ میانی، مقادیر تصادفی اولیه به شکل $c_r^{j,(0)}$ و $\sigma_r^{j,(0)}$ و در بازهٔ [0 1] نسبت می‌دهیم. برای این دو پارامتر یعنی $c_r^{j,(t)}$ و $\sigma_r^{j,(t)}$، اندیس‌های j، r و t همانند گذشته تعریف می‌گردند.

۲. بردار ورودی آموزشی جدید را اعمال می‌کنیم.

$$X^{(t)} + \mathcal{N}^{(t)} = [(x_1^{(t)} + n_1^{(t)}) \;\; (x_2^{(t)} + n_2^{(t)}) \dots (x_n^{(t)} + n_n^{(t)})] \quad (۱۴)$$

[1] Unsupervisory Learning
[2] Supervisory Learning

۳. به یکی از دو روش زیر، تعلق بردار ورودی آموزشی جدید $X^{(t)}+\mathcal{N}^{(t)}$ را به یک خوشه تعیین می‌کنیم.

۳. ۱. نزدیکترین مرکزِ توابعِ فعال‌سازِ گاوسی، نسبت به بردار $X^{(t)}+\mathcal{N}^{(t)}$ را به دست می‌آوریم. در این‌جا از فاصله اقلیدسی به عنوان معیار نزدیکی بین بردار ورودی $X^{(t)}+\mathcal{N}^{(t)}$ و مرکز خوشه $c^{j,(t)}$ به صورت زیر استفاده می‌شود.

$$J_j^{(t)}=\sqrt{(\|X^{(t)}+\mathcal{N}^{(t)}-c^{j,(t)}\|_2)}=\sqrt{\sum_{r=1}^{n}\left(x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}\right)^2} \tag{۱۵}$$

۳. ۲. تعلق «هر بردار ورودی به یک خوشه» را با استفاده از ماتریس تعلق[1] U به ابعاد $n\times m$ تعیین می‌کنیم. در این ماتریس دودویی، درایه $u_r^{j,(t)}$ برابر یک است اگر rاُمین داده بردار ورودی یعنی $x_r^{(t)}+n_r^{(t)}$ به گروه j تعلق داشته باشد و در غیر این‌صورت برابر صفر خواهد بود. این ماتریس می‌تواند به‌صورت زیر توصیف شود:

$$U^{(t)}=\begin{bmatrix} u_1^{1,(t)} & u_1^{2,(t)} & \dots & u_1^{m,(t)} \\ u_2^{1,(t)} & u_2^{2,(t)} & \dots & u_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ u_n^{1,(t)} & u_n^{2,(t)} & \dots & u_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۱۶}$$

که در آن داریم:

$$u_r^{j,(t)}=\begin{cases}1 & if\ \left\|x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}\right\|_2\le\left\|x_r^{(t)}+n_r^{(t)}-c_r^{k,(t)}\right\|_2\\ 0 & otherwise\end{cases}$$
$$\forall\, k=1,2,\dots,m,\ \forall r=1,2,3,\dots,n \mid k\neq j \tag{۱۷}$$

ماتریس تعلق دودویی U لازم است که هر دو ویژگی (۱۸) و (۱۹) را داشته باشد.

$$\sum_{j=1}^{m}u_r^{j,(t)}=1 \qquad \forall r=1,2,3,\dots,n \tag{۱۸}$$
$$\sum_{r=1}^{n}\sum_{j=1}^{m}u_r^{j,(t)}=n \tag{۱۹}$$

از آنجایی‌که هر بردار ورودی تنها می‌تواند به یک خوشه تعلق داشته باشد، با ضرب بردار ورودی آموزشی جدید $X^{(t)}+\mathcal{N}^{(t)}$ در ماتریس تعلق U به ماتریس زیر می‌رسیم:

$$\left(X^{(t)}+\mathcal{N}^{(t)}\right)\times U^{(t)}=\left[\left(x_1^{(t)}+n_1^{(t)}\right)\ \left(x_2^{(t)}+n_2^{(t)}\right)\dots\left(x_n^{(t)}+n_n^{(t)}\right)\right]_{1\times n}\times\begin{bmatrix} u_1^{1,(t)} & u_1^{2,(t)} & \dots & u_1^{m,(t)} \\ u_2^{1,(t)} & u_2^{2,(t)} & \dots & u_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ u_n^{1,(t)} & u_n^{2,(t)} & \dots & u_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۲۰}$$

$$\left(X^{(t)}+\mathcal{N}^{(t)}\right)\times U^{(t)}=\Big[\left(x_1^{(t)}+n_1^{(t)}\right).u_1^{1,(t)}+\left(x_2^{(t)}+n_2^{(t)}\right).u_2^{1,(t)}+\dots+\left(x_n^{(t)}+n_n^{(t)}\right).u_n^{1,(t)}$$
$$\left(x_1^{(t)}+n_1^{(t)}\right).u_1^{2,(t)}+\left(x_2^{(t)}+n_2^{(t)}\right).u_2^{2,(t)}+\dots+\left(x_n^{(t)}+n_n^{(t)}\right).u_n^{2,(t)}\quad\dots$$
$$\left(x_1^{(t)}+n_1^{(t)}\right).u_1^{m,(t)}+\left(x_2^{(t)}+n_2^{(t)}\right).u_2^{m,(t)}+\dots+\left(x_n^{(t)}+n_n^{(t)}\right).u_n^{m,(t)}\Big] \tag{۲۱}$$

هر درایه ماتریس حاصلضرب مربوط به یک خوشه است و هر چه مقدار یک درایه بزرگتر باشد احتمال تعلق بردار ورودی به آن خوشه بیشتر می‌گردد. قطعی‌ترین حالت زمانی است که به جز یک درایه بقیه آن‌ها صفر باشند و بدترین وضعیت زمانی رخ می‌دهد که مقدار دو یا چند درایه مساوی گردد که در این صورت تعلق بردار ورودی به یکی از خوشه‌هایی که مقدار درایه مربوط به آن‌ها مساوی است، به صورت تصادفی تعیین می‌گردد.

۴. تابع هزینه می تواند توسط رابطه زیر تعریف گردد:

$$J^{(t)}=\sum_{j=1}^{m}J_j^{(t)}=\sum_{j=1}^{m}\left\|X^{(t)}+\mathcal{N}^{(t)}-c^{j,(t)}\right\|_2=\sum_{j=1}^{m}\sqrt{\sum_{r=1}^{n}\left(x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}\right)^2} \tag{۲۲}$$

۵. مرکز خوشه $c^{j,(t)}$ را به کمک رابطهٔ (۲۳) و انحراف استاندارد خوشه $\sigma^{j,(t)}$ را به وسیلهٔ رابطهٔ (۲۴) تغییر می‌دهیم.

$$c^{j,(t+1)}=c^{j,(t)}+\lambda_c^{(t)}(t)(X^{(t)}+\mathcal{N}^{(t)}-c^{j,(t)}) \tag{۲۳}$$
$$\sigma^{j,(t+1)}=\sigma^{j,(t)}+\lambda_\sigma^{(t)}(t)(X^{(t)}+\mathcal{N}^{(t)}-\sigma^{j,(t)}) \tag{۲۴}$$

پارامترهای $\lambda_c^{(t)}$ و $\lambda_\sigma^{(t)}$ ضرایب آموزش تطبیقی هستند که توسط یک تابع خطی از t به شکل زیر آموزش می‌بینند.

$$\lambda_c^{(t)}=\lambda_c^{(0)}(1-\frac{t}{T}) \tag{۲۵}$$
$$\lambda_\sigma^{(t)}=\lambda_\sigma^{(0)}(1-\frac{t}{T}) \tag{۲۶}$$

پارامترهای $\lambda_c^{(0)}$ و $\lambda_\sigma^{(0)}$ ضرایب آموزش تطبیقی اولیه، پارامتر t دورهٔ آموزشی فعلی و T تعداد کل تکرار برای آموزش است.

۶. شرط خاتمه حلقه[2] می‌تواند به یکی از سه شکل زیر در نظر گرفته شود که در صورت برقراری یکی از آن‌ها آموزش پایان می‌یابد و در غیر اینصورت به گام دو برمی‌گردیم.

الف) بعد از طی تعداد تعیین شده‌ای از تکرار.

ب) در صورتی‌که برای تابع هزینه مقدار معین بدست آمده باشد.

ج) بهبود تابع هزینه نسبت به تکرار قبلی کمتر از یک حد آستانه معین باشد.

[1] Membership matrix

[2] End of loop condition

۳-۱-۱ الگوریتم پس‌انتشار خطا با استفاده از گرادیان نزولی[1]

الگوریتم پس انتشار خطا برای n ورودی و m خوشه و یک خروجی، بر اساس مجموع مربعاتِ خطا به صورت زیر محاسبه می‌گردد در این الگوریتم نرخ آموزش می‌تواند برای تمام پارامترها یکسان یا متفاوت در نظر گرفته شود[۳۳ ۳۵].

$$e^{(t)} = d^{(t)} - y_{\mathcal{N}}{}^{(t)}\ , \qquad E^{(t)} = \frac{1}{2}\left(e^{(t)}\right)^2 \tag{۲۷}$$

برای وزن‌های لایهٔ خروجی که پارامترهای خطی می‌باشند داریم:

$$\Delta v^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial v^{j,(t)}} = -\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial y^{(t)}} \times \frac{\partial y^{(t)}}{\partial v^{j,(t)}} \tag{۲۸}$$

$$\Delta v^{j,(t)} = -\eta \times e^{(t)} \times (-1) \times o^{j,(t)} \tag{۲۹}$$

$$v^{j,(t+1)} = v^{j,(t)} + \Delta v^{j,(t)} = v^{j,(t)} + \eta e^{(t)} o^{j,(t)} \tag{۳۰}$$

برای پارامترهای لایهٔ میانی خواهیم داشت:

$$\Delta c_r^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial c_r^{j,(t)}} = -\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial y^{(t)}} \times \frac{\partial y^{(t)}}{\partial o^{j,(t)}} \times \frac{\partial o^{j,(t)}}{\partial d^{j,(t)}} \times \frac{\partial d^{j,(t)}}{\partial c_r^{j,(t)}} \tag{۳۱}$$

$$\Delta c_r^{j,(t)} = -\eta \times e^{(t)} \times (-1) \times v^{j,(t)} \times \frac{o^{j,(t)} \times d^{j,(t)}}{x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}} \tag{۳۲}$$

$$c_r^{j,(t+1)} = c_r^{j,(t)} + \Delta c_r^{j,(t)} = c_r^{j,(t)} + \eta e^{(t)} v^{j,(t)} \frac{o^{j,(t)} \times d^{j,(t)}}{x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}} \tag{۳۳}$$

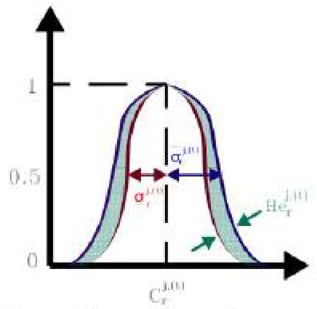
$$\Delta \sigma_r^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial \sigma_r^{j,(t)}} = -\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial y^{(t)}} \times \frac{\partial y^{(t)}}{\partial o^{j,(t)}} \times \frac{\partial o^{j,(t)}}{\partial d^{j,(t)}} \times \frac{\partial d^{j,(t)}}{\partial \sigma_r^{j,(t)}} \tag{۳۴}$$

$$\Delta \sigma_r^{j,(t)} = -\eta \times e^{(t)} \times (-1) \times v^{j,(t)} \times \frac{o^{j,(t)} \times d^{j,(t)}}{\sigma_r^{j,(t)}} \tag{۳۵}$$

$$\sigma_r^{j,(t+1)} = \sigma_r^{j,(t)} + \Delta \sigma_r^{j,(t)} = \sigma_r^{j,(t)} + \eta e^{(t)} v^j \frac{d^{j,(t)} \times o^{j,(t)}}{\sigma_r^{j,(t)}} \tag{۳۶}$$

۴- معرفی شبکهٔ عصبی بر پایه تابعِ فعال‌سازِ گاوسیِ گرانولی

نمونه‌ای از تابع فعال‌ساز گاوسیِ گرانولی بکار رفته در نُرون‌های لایهٔ میانی، در شکل۳ نشان داده شده است. به این نوع توابع، تابع گاوسیِ مدل ابر[2] گفته می‌شود و دارای سه مشخصهٔ عددی می‌باشند. این سه مشخصهٔ عددی ۱) مرکز تابع گاوسی ۲) عدم قطعیت اندازه‌گیری شده و ۳) بالاترین مقدار عدم قطعیت هستند که به ترتیب و در ادامهٔ مطلب تعریف شده‌اند[۱۷، ۲۱، ۲۳].



شکل۳. تابع فعال‌ساز گاوسیِ گرانولی در نُرون‌های لایهٔ میانی

۱. نمونه‌های دارای بیشترین تکرار را با $c_r^{j,(t)}$ نشان می دهند که مرکز یک تابع گاوسی است. در پیاده سازی، مراکز توابع گاوسی در یک ماتریس به نام $(mean^{(t)})^T$ ذخیره می‌گردند. تعداد سطرهای این ماتریس همان تعداد ورودی‌های شبکهٔ عصبی و تعداد ستون‌های آن برابر تعداد نُرون‌های لایهٔ میانی است. این مقادیر بازه‌ای نیستند[۱۸، ۲۳].

$$\left(\boldsymbol{mean}^{(t)}\right)^{\boldsymbol{T}} = \boldsymbol{C}^{(t)} = \left[\boldsymbol{c}^{1,(t)} \quad \boldsymbol{c}^{2,(t)} \quad \dots \quad \boldsymbol{c}^{m,(t)}\right]_{1\times m} = \begin{bmatrix} c_1^{1,(t)} & c_1^{2,(t)} & \dots & c_1^{m,(t)} \\ c_2^{1,(t)} & c_2^{2,(t)} & \dots & c_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ c_n^{1,(t)} & c_n^{2,(t)} & \dots & c_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۳۷}$$

۲. عدم قطعیت اندازه‌گیری شده را با متغیر إنتروپی[3] [۲۳] مُدل می‌کنند که در ارتباط با انحراف استاندارد است. انحرافِ استاندارد دارای مقادیر بازه‌ای است و کرانِ پائینِ آن در یک ماتریس به نام $(STDEVleft^{(t)})^T$ و کرانِ بالا در ماتریس دیگری به نام $(STDEVright^{(t)})^T$ ذخیره می‌گردد. در هر دو ماتریس تعداد سطرها همان تعداد ورودی‌های شبکهٔ عصبی و تعداد ستون‌ها تعداد نُرون‌های لایهٔ میانی است[۱۸، ۲۶]. هر چه میزان نویز تزریق شده به داده‌ها بیشتر باشد فاصله کرانِ پائینِ انحرافِ استاندارد از کرانِ بالای آن بیشتر می‌گردد و در نتیجه إنتروپی بزرگتری خواهیم داشت که با پهن‌تر شدن تابع فعال‌سازِ گاوسیِ گرانولی برای مدیریت عدم قطعیت بزرگتر همراه است[۴، ۵].

$$\left(\boldsymbol{STDEVleft}^{(t)}\right)^{\boldsymbol{T}} = \underline{\boldsymbol{\Sigma}}^{(t)} = \left[\underline{\boldsymbol{\sigma}}^{1,(t)} \quad \underline{\boldsymbol{\sigma}}^{2,(t)} \quad \dots \quad \underline{\boldsymbol{\sigma}}^{m,(t)}\right]_{1\times m} = \begin{bmatrix} \underline{\sigma}_1^{1,(t)} & \underline{\sigma}_1^{2,(t)} & \dots & \underline{\sigma}_1^{m,(t)} \\ \underline{\sigma}_2^{1,(t)} & \underline{\sigma}_2^{2,(t)} & \dots & \underline{\sigma}_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\sigma}_n^{1,(t)} & \underline{\sigma}_n^{2,(t)} & \dots & \underline{\sigma}_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۳۸}$$

[1] Error Back Propagation with Steepest Gradient Descent
[2] Cloud model
[3] Entropy (En)

$$\left(STDEVright^{(t)}\right)^T = \overline{\Sigma}^{(t)} = [\overline{\sigma}^{1,(t)} \quad \overline{\sigma}^{2,(t)} \quad \dots \quad \overline{\sigma}^{m,(t)}]_{1\times m} = \begin{bmatrix} \overline{\sigma}_1^{1,(t)} & \overline{\sigma}_1^{2,(t)} & \dots & \overline{\sigma}_1^{m,(t)} \\ \overline{\sigma}_2^{1,(t)} & \overline{\sigma}_2^{2,(t)} & \dots & \overline{\sigma}_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\sigma}_n^{1,(t)} & \overline{\sigma}_n^{2,(t)} & \dots & \overline{\sigma}_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۳۹}$$

$$Entropy^{(t)} = \overline{\Sigma}^{(t)} - \underline{\Sigma}^{(t)} = [\overline{\sigma}^{1,(t)} \quad \overline{\sigma}^{2,(t)} \quad \dots \quad \overline{\sigma}^{m,(t)}]_{1\times m} - [\underline{\sigma}^{1,(t)} \quad \underline{\sigma}^{2,(t)} \quad \dots \quad \underline{\sigma}^{m,(t)}]_{1\times m} \tag{۴۰}$$

$$Entropy^{(t)} = \begin{bmatrix} \overline{\sigma}_1^{1,(t)} & \overline{\sigma}_1^{2,(t)} & \dots & \overline{\sigma}_1^{m,(t)} \\ \overline{\sigma}_2^{1,(t)} & \overline{\sigma}_2^{2,(t)} & \dots & \overline{\sigma}_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\sigma}_n^{1,(t)} & \overline{\sigma}_n^{2,(t)} & \dots & \overline{\sigma}_n^{m,(t)} \end{bmatrix}_{n\times m} - \begin{bmatrix} \underline{\sigma}_1^{1,(t)} & \underline{\sigma}_1^{2,(t)} & \dots & \underline{\sigma}_1^{m,(t)} \\ \underline{\sigma}_2^{1,(t)} & \underline{\sigma}_2^{2,(t)} & \dots & \underline{\sigma}_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\sigma}_n^{1,(t)} & \underline{\sigma}_n^{2,(t)} & \dots & \underline{\sigma}_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۴۱}$$

$$Entropy^{(t)} = \begin{bmatrix} \overline{\sigma}_1^{1,(t)} - \underline{\sigma}_1^{1,(t)} & \overline{\sigma}_1^{2,(t)} - \underline{\sigma}_1^{2,(t)} & \dots & \overline{\sigma}_1^{m,(t)} - \underline{\sigma}_1^{m,(t)} \\ \overline{\sigma}_2^{1,(t)} - \underline{\sigma}_2^{1,(t)} & \overline{\sigma}_2^{2,(t)} - \underline{\sigma}_2^{2,(t)} & \dots & \overline{\sigma}_2^{m,(t)} - \underline{\sigma}_2^{m,(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\sigma}_n^{1,(t)} - \underline{\sigma}_n^{1,(t)} & \overline{\sigma}_n^{2,(t)} - \underline{\sigma}_n^{2,(t)} & \dots & \overline{\sigma}_n^{m,(t)} - \underline{\sigma}_n^{m,(t)} \end{bmatrix}_{n\times m} \tag{۴۲}$$

۳. بالاترین مقدار عدم قطعیت را با هایپر اِنتروپی[1] [۲۳] نشان می‌دهند. هنگامی که کران‌های پائین و بالای انحرافِ استاندارد یعنی مقادیر $\underline{\sigma}_r^{j,(t)}$ و $\overline{\sigma}_r^{j,(t)}$ برای هر تابع فعال‌ساز گرانولی آموزش داده می‌شود، مقدار He تغییر می‌کند. البته مقادیر $\underline{\sigma}_r^{j,(t)}$ و $\overline{\sigma}_r^{j,(t)}$ در تابع فعال‌ساز گرانولی، نسبت به خط افقی گذرنده از وسط محور عمودی سنجیده می‌شوند. مقدار He طبق شکل۲ در جایی بوجود می‌آید که تابع فعال‌ساز گرانولی بهترین موقعیت خود را دارد. منطقی است که با مقدار قوی‌تر نویز در مقادیر ورودی و پس از آموزش کران‌های پائین و بالای انحرافِ استاندارد، مقدار He بزرگتری بدست آید و برعکس مقدار ضعیف‌تر نویز مقدار He کوچکتری را نتیجه دهد. در واقع بازه‌ای بودن انحرافِ استاندارد و وزن‌های لایهٔ خروجی، عدم قطعیت حاصل از نویز در داده‌های ورودی را پوشش داده و ابزار غلبه بر آن‌ها را در اختیار شبکهٔ عصبیِ RBF گرانولی قرار می‌دهد[۶ ۱۰].

همانطور که گفته شد، مقادیر مرکزِ توابعِ فعال‌سازِ گاوسیِ گرانولیِ نُرون زاُم لایهٔ میانی، در یک بردار به نام $c^{j,(t)}$ ذخیره می‌گردد که تعداد درایه‌های آن برابر تعداد ورودی‌های سیستم می‌باشد. درایه‌های این ماتریس مقادیر ثابت بوده و بازه‌ای نیستند. مقادیر انحرافِ استاندارد که نامعین می‌باشند و به صورت بازهٔ $[\, \underline{\sigma}_r^{j,(t)} \quad \overline{\sigma}_r^{j,(t)} \,]$ در نظر گرفته می‌شود نیز در دو ماتریس کرانِ پائین و کرانِ بالا نگهداری می‌گردند.

یک شبکهٔ عصبیِ RBF گرانولی با استفاده از تابعِ گاوسیِ مُدلِ ابر، یک شبکهٔ عصبیِ گرانولی با مقادیر انحرافِ استانداردِ غیرقطعی[2] هم نامیده می‌شود[۵]. خروجی هر نُرون لایهٔ میانی در این شبکهٔ عصبیِ RBF گرانولی دارای کرانِ پائین و کرانِ بالا بوده و آن را می‌توان به صورت یک بازهٔ $[\, \underline{o}^{j,(t)} \quad \overline{o}^{j,(t)} \,]$ نشان داد. ماتریس‌های کرانِ پائین و کرانِ بالایِ خروجیِ لایهٔ پنهان شبکهٔ عصبیِ RBF گرانولی برای یک دسته دادهٔ ورودی برابر خواهد بود با[۲۳]:

$$\underline{\boldsymbol{O}}^{(t)} = [\underline{o}^{1,(t)} \quad \underline{o}^{2,(t)} \dots \underline{o}^{m,(t)}] = [\psi(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{1,(t)}, \underline{\boldsymbol{\sigma}}^{1,(t)}) \quad \psi(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{2,(t)}, \underline{\boldsymbol{\sigma}}^{2,(t)}) \dots \dots \psi(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{m,(t)}, \underline{\boldsymbol{\sigma}}^{m,(t)})] \tag{۴۳}$$

$$\underline{\boldsymbol{O}}^{(t)} = \left[\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{1,(t)}}{\underline{\sigma}_r^{1,(t)}}\right)^2\right] \quad \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{2,(t)}}{\underline{\sigma}_r^{2,(t)}}\right)^2\right] \dots \cdots \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{m,(t)}}{\underline{\sigma}_r^{m,(t)}}\right)^2\right]\right] \tag{۴۴}$$

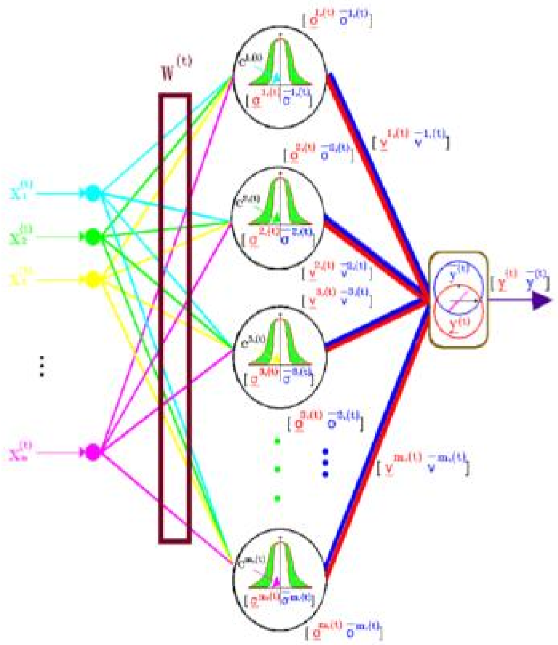
$$\overline{\boldsymbol{O}}^{(t)} = [\overline{o}^{1,(t)} \quad \overline{o}^{2,(t)} \dots \overline{o}^{m,(t)}] = \left[\psi\left(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{1,(t)}, \overline{\boldsymbol{\sigma}}^{1,(t)}\right) \quad \psi\left(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{2,(t)}, \overline{\boldsymbol{\sigma}}^{2,(t)}\right) \dots \dots \psi\left(\boldsymbol{X}^{(t)}, \boldsymbol{\mathcal{N}}^{(t)}, \boldsymbol{c}^{m,(t)}, \overline{\boldsymbol{\sigma}}^{m,(t)}\right)\right] \tag{۴۵}$$

$$\overline{\boldsymbol{O}}^{(t)} = \left[\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{1,(t)}}{\overline{\sigma}_r^{1,(t)}}\right)^2\right] \quad \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{2,(t)}}{\overline{\sigma}_r^{2,(t)}}\right)^2\right] \dots \cdots \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)} + n_r^{(t)} - c_r^{m,(t)}}{\overline{\sigma}_r^{m,(t)}}\right)^2\right]\right] \tag{۴۶}$$

در شکلِ۴ یک شبکهٔ عصبیِ RBF گرانولی که بر پایه توابعِ فعال‌سازِ گرانولی بنا گردیده، نشان داده شده است. در این شبکهٔ عصبیِ؛ انحرافِ استاندارد بازه‌ای در توابعِ فعال‌سازِ گرانولیِ لایهٔ پنهان و وزن‌هایِ بازه‌ای در لایهٔ خروجی؛ انعطاف‌پذیری و پایداری آن در مقابل نویز را به مقدار زیادی افزایش می‌دهند.

[1] Hyper Entropy (He)

[2] Uncertain standard deviation



شکل ۴. شبکهٔ عصبی RBF گرانولی بر پایه تابع فعال‌ساز گرانولی

محاسبه کرانِ پائینِ خروجیِ یک تابع گاوسیِ گرانولی در لایهٔ میانی به شکل زیر است [۴، ۵، ۳۷]:

$$\underline{o}^{j,(t)} = \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}}{\underline{\sigma}_r^{j,(t)}}\right)^2\right] = \psi\left[\sum_{r=1}^{n}\left(\overline{u}_r^{j,(t)}\right)^2\right] = \psi\left(\overline{d}^{j,(t)}\right) = exp\left(-\frac{1}{2}\overline{d}^{j,(t)}\right) \quad (۴۷)$$

وجود کرانِ پائینِ انحرافِ استاندارد، به ازای «ورودی r اُم به نُرون jاُم لایهٔ پنهانِ شبکه، یعنی $\underline{\sigma}_r^{j,(t)}$، در مخرجِ کسرِ $\frac{x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}}{\underline{\sigma}_r^{j,(t)}}$ سبب بزرگ شدن آن و در نتیجه محاسبه $\overline{u}_r^{j,(t)}$ خواهد شد. با به توان دو رساندن و جمع کردن مقادیر $\overline{u}_r^{j,(t)}$ به ازای تمام ورودی‌هایِ $x_r^{(t)} + n_r^{(t)}$ با ابعاد $r = 1, 2, \dots, n$ در نُرون jاُم لایهٔ پنهان، مقدار مثبت $\overline{d}^{j,(t)}$ به دست می‌آید. هنگامی که $\overline{d}^{j,(t)}$ در ضریبِ $-\frac{1}{2}$ ضرب واز حاصل‌ضرب تابع نمایی به صورت $exp\left(-\frac{1}{2}\overline{d}^{j,(t)}\right) = \frac{1}{e^{\left(\frac{1}{2}\overline{d}^{j,(t)}\right)}}$ گرفته می‌شود، به دلیل بزرگ شدن مخرج کسر، حاصل تقسیم کوچک شده و کرانِ پائینِ خروجیِ تابع گاوسیِ گرانولی یعنی $\underline{o}^{j,(t)}$ که مقداری مثبت است، به دست می‌آید.

اکنون برای محاسبه کرانِ بالایِ خروجیِ یک تابع گاوسیِ گرانولی خواهیم داشت[۴، ۵، ۳۸]:

$$\overline{o}^{j,(t)} = \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}}{\overline{\sigma}_r^{j,(t)}}\right)^2\right] = \psi\left[\sum_{r=1}^{n}\left(\underline{u}_r^{j,(t)}\right)^2\right] = \psi\left(\underline{d}^{j,(t)}\right) = exp\left(-\frac{1}{2}\underline{d}^{j,(t)}\right) \quad (۴۸)$$

وجود کرانِ بالایِ انحرافِ استاندارد، به ازای «ورودی r اُم به نُرون jاُم لایهٔ پنهانِ شبکه، یعنی $\overline{\sigma}_r^{j,(t)}$، در مخرجِ کسرِ $\frac{x_r^{(t)}+n_r^{(t)}-c_r^{j,(t)}}{\overline{\sigma}_r^{j,(t)}}$ سبب کوچک شدن آن و در نتیجه محاسبه $\underline{u}_r^{j,(t)}$ خواهد شد. با به توان دو رساندن و جمع کردن مقادیر $\underline{u}_r^{j,(t)}$ به ازای تمام ورودی‌هایِ $x_r^{(t)} + n_r^{(t)}$ با ابعاد $r = 1, 2, \dots, n$ در نُرون jاُم لایهٔ پنهان، مقدار مثبت $\underline{d}^{j,(t)}$ به دست می‌آید. هنگامی که $\underline{d}^{j,(t)}$ در ضریبِ $-\frac{1}{2}$ ضرب واز حاصل‌ضرب تابع نمایی به صورت $exp\left(-\frac{1}{2}\underline{d}^{j,(t)}\right) = \frac{1}{e^{\left(\frac{1}{2}\underline{d}^{j,(t)}\right)}}$ گرفته می‌شود، به دلیل کوچک شدن مخرج کسر، حاصل تقسیم بزرگ شده و کرانِ بالایِ خروجیِ تابع گاوسیِ گرانولی یعنی $\overline{o}^{j,(t)}$ که مقداری مثبت است، به دست می‌آید.

همانطور که مشاهده می‌گردد گاوسی بودن یک تابع فعال‌سازِ گرانولی تضمین می‌کند که کرانِ پائین و کرانِ بالایِ خروجیِ آن همیشه مثبت باشد. پس هر مقدار متعلق به بازهٔ $[\underline{o}^{j,(t)} \quad \overline{o}^{j,(t)}]$، مقداری مثبت خواهد بود.

از طرف دیگر با توجه به این‌که $s^{j,(t)}$ بازهٔ تغییرات «مرکزِ وزنِ اتصالِ بین لایهٔ میانی و لایهٔ خروجی برای jاُمین نُرون لایهٔ پنهان» یعنی $v^{j,(t)}$ می‌باشد. و از آنجا که در محاسبه کرانِ پائین و کرانِ بالایِ «وزن‌هایِ بازه‌ای در لایهٔ خروجی»، یعنی مقادیر $\underline{v}^{j,(t)}$ و $\overline{v}^{j,(t)}$ از «قدرمطلق بازهٔ این وزن‌ها» یعنی $|s^{j,(t)}|$، به صورت $\underline{v}^{j,(t)} = v^{j,(t)} - |s^{j,(t)}|$ و $\overline{v}^{j,(t)} = v^{j,(t)} + |s^{j,(t)}|$ استفاده گردیده است، تضمین می‌گردد که به ازای مقادیر منفی و مثبتِ «مرکزِ وزن‌هایِ بازه‌ای در لایهٔ خروجی»، یعنی $v^{j,(t)}$ کرانِ پائینِ «وزن‌هایِ بازه‌ای در لایهٔ خروجی» یعنی مقدار $\underline{v}^{j,(t)}$ همیشه کوچکتر از کرانِ بالایِ «وزن‌هایِ بازه‌ای در لایهٔ خروجی» یعنی مقدار $\overline{v}^{j,(t)}$ باشد. یعنی همیشه رابطه نامساوی به صورت $\underline{v}^{j,(t)} < \overline{v}^{j,(t)}$ که معادل با رابطه نامساوی به شکل $v^{j,(t)} - |s^{j,(t)}| < v^{j,(t)} + |s^{j,(t)}|$ است، برقرار می‌باشد.

سرانجام، برای محاسبهٔ خروجیِ نهاییِ شبکهٔ عصبیِ RBF گرانولی چهار حالت زیر ممکن است رخ دهد.

- حالت اوّل: اگر $v^{j,(t)} > 0$ و $v^{j,(t)} > |s^{j,(t)}|$ باشد آن‌گاه $\underline{v}^{j,(t)}, \overline{v}^{j,(t)} > 0$ خواهد بود.
- حالت دوّم: اگر $v^{j,(t)} > 0$ و $v^{j,(t)} < |s^{j,(t)}|$ باشد آن‌گاه $\underline{v}^{j,(t)} < 0$ و $\overline{v}^{j,(t)} > 0$ خواهد بود.
- حالت سوّم: اگر $v^{j,(t)} < 0$ و $|v^{j,(t)}| > |s^{j,(t)}|$ باشد آن‌گاه $\underline{v}^{j,(t)}, \overline{v}^{j,(t)} < 0$ خواهد بود.
- حالت چهارم: اگر $v^{j,(t)} < 0$ و $|v^{j,(t)}| < |s^{j,(t)}|$ باشد آن‌گاه $\underline{v}^{j,(t)} < 0$ و $\overline{v}^{j,(t)} > 0$ خواهد بود.

در حالت‌های اوّل، دوّم و چهارم، کران پایین و کران بالایِ خروجیِ نهاییِ شبکهٔ عصبیِ RBF گرانولی، به همراه پارامترهای خطی کران پایین و کران بالایِ لایهٔ خروجیِ شبکه به این صورت خواهد بود[۲۳]:

$$\underline{y_{\mathcal{N}}}^{(t)} = \sum_{j=1}^{m}\left(\underline{v}^{j,(t)} \times \underline{o}^{j,(t)}\right) = \sum_{j=1}^{m}\left[\left(v^{j,(t)} - \left|s^{j,(t)}\right|\right) \times \underline{o}^{j,(t)}\right] \quad (۴۹)$$

$$\overline{y_{\mathcal{N}}}^{(t)} = \sum_{j=1}^{m}\left(\overline{v}^{j,(t)} \times \overline{o}^{j,(t)}\right) = \sum_{j=1}^{m}\left[\left(v^{j,(t)} + \left|s^{j,(t)}\right|\right) \times \overline{o}^{j,(t)}\right] \quad (۵۰)$$

در حالت سوّم به علت منفی بودن $\underline{v}^{j,(t)}$ و $\overline{v}^{j,(t)}$ و مثبت بودن $\underline{o}^{j,(t)}$ و $\overline{o}^{j,(t)}$ محاسبه کران پایین و کران بالای خروجی نهایی شبکهٔ عصبی RBF گرانولی، به همراه پارامترهای خطی کران پایین و کران بالای لایهٔ خروجی شبکه به این صورت خواهد بود:

$$\underline{y_{\mathcal{N}}}^{(t)} = \sum_{j=1}^{m}\left(\underline{v}^{j,(t)} \times \overline{o}^{j,(t)}\right) = \sum_{j=1}^{m}\left[\left(v^{j,(t)} - \left|s^{j,(t)}\right|\right) \times \overline{o}^{j,(t)}\right] \quad (۵۱)$$

$$\overline{y_{\mathcal{N}}}^{(t)} = \sum_{j=1}^{m}\left(\overline{v}^{j,(t)} \times \underline{o}^{j,(t)}\right) = \sum_{j=1}^{m}\left[\left(v^{j,(t)} + \left|s^{j,(t)}\right|\right) \times \underline{o}^{j,(t)}\right] \quad (۵۲)$$

و خروجی نهایی از میانگین کران پائین و کران بالا از رابطه (۵۲) و یا به فرم تطبیق‌پذیر[1] از رابطه (۵۳) به دست می‌آید[۶]:

$$y_{\mathcal{N}}^{(t)} = \frac{\underline{y_{\mathcal{N}}}^{(t)} + \overline{y_{\mathcal{N}}}^{(t)}}{2} \quad (۵۳)$$

$$y_{\mathcal{N}}^{(t)} = \alpha_l \times \underline{y_{\mathcal{N}}}^{(t)} + \alpha_u \times \overline{y_{\mathcal{N}}}^{(t)} \quad (۵۴)$$

که در آن α_l و α_u به ترتیب پارامترهای تطبیقی کران پائین و کران بالای خروجی هستند.

۴-۱ الگوریتم خوشه‌بندی K-Means با m خوشه

گام‌های الگوریتم برای آموزش پارامترهای لایهٔ میانی شبکهٔ عصبی RBF گرانولی، کاملاً مشابه الگوریتم آموزش برای نوع غیربازه‌ای این شبکهٔ عصبی می‌باشد. تنها تفاوت در این است که در این‌جا هر دو کران پائین و کران بالا در «انحراف استاندارد توابع فعال‌ساز گرانولی» به طور جداگانه آموزش می‌بینند. پس به جای دو پارامتر، سه پارامتر برای آموزش وجود دارد[۲۳، ۳۷].

۴-۲ الگوریتم پس‌انتشار خطای گرانولی[۳۳، ۳۵] با استفاده از گرادیان نزولی[2]

الگوریتم پس‌انتشار خطای گرانولی برای n ورودی و m خوشه و یک خروجی، بر اساس روش مجموع مربعات خطا به صورت زیر محاسبه می‌گردد[۴، ۵، ۳۸].

$$e^{(t)} = d^{(t)} - y_{\mathcal{N}}^{(t)}, \qquad E^{(t)} = \frac{1}{2}\left(e^{(t)}\right)^2, \qquad e^{(t)} = d^{(t)} - \left(\frac{\underline{y_{\mathcal{N}}}^{(t)}}{2} + \frac{\overline{y_{\mathcal{N}}}^{(t)}}{2}\right) \quad (۵۵)$$

وزن‌های لایهٔ خروجی را به این صورت آموزش می‌دهیم که برای هر بازه یک «مقدار مرکزی»[3] و یک «مقدار تغییرات»[4] که می‌تواند در دو طرف مقدار مرکزی یکسان باشد[5] و یا یکسان نباشد[6]، در نظر می‌گیریم. پس دو پارامتر در لایهٔ خروجی آموزش می‌بینند که الف) مرکز وزن‌های بازه‌ای و ب) بازهٔ وزن‌ها می‌باشند. برای مرکزِ وزن‌های لایهٔ خروجی داریم:

$$\Delta v^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial v^{j,(t)}} = \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \underline{y}^{(t)}} \times \frac{\partial \underline{y}^{(t)}}{\partial v^{j,(t)}}\right) + \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \overline{y}^{(t)}} \times \frac{\partial \overline{y}^{(t)}}{\partial v^{j,(t)}}\right) \quad (۵۶)$$

$$\Delta v^{j,(t)} = \left[-\eta \times e^{(t)} \times \left(-\frac{1}{2}\right) \times \underline{o}^{j,(t)}\right] + \left[-\eta \times e^{(t)} \times \left(-\frac{1}{2}\right) \times \overline{o}^{j,(t)}\right] \quad (۵۷)$$

$$\Delta v^{j,(t)} = \left(\frac{1}{2} \times \eta \times e^{(t)} \times \underline{o}^{j,(t)}\right) + \left(\frac{1}{2} \times \eta \times e^{(t)} \times \overline{o}^{j,(t)}\right) \quad (۵۸)$$

$$v^{j,(t+1)} = v^{j,(t)} + \Delta v^{j,(t)} = v^{j,(t)} + \eta e^{(t)} o^{j,(t)} \quad (۵۹)$$

و برای بازهٔ وزن‌های لایهٔ خروجی داریم:

[1] Adaptive form

[2] Granular Error Back Propagation with Steepest Gradient Descent
[3] Midpoint
[4] Interval
[5] Symmetrical
[6] Asymmetrical

$$\Delta s^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial s^{j,(t)}} = \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \underline{y}^{(t)}} \times \frac{\partial \underline{y}^{(t)}}{\partial s^{j,(t)}}\right) + \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \overline{y}^{(t)}} \times \frac{\partial \overline{y}^{(t)}}{\partial s^{j,(t)}}\right) \qquad (۶۰)$$

$$\Delta s^{j,(t)} = \left[-\eta \times e^{(t)} \times \left(-\tfrac{1}{2}\right) \times \underline{o}^{j,(t)}\right] + \left[-\eta \times e^{(t)} \times \left(-\tfrac{1}{2}\right) \times \overline{o}^{j,(t)}\right] \qquad (۶۱)$$

$$\Delta s^{j,(t)} = \left(\tfrac{1}{2} \times \eta \times e^{(t)} \times \underline{o}^{j,(t)}\right) + \left(\tfrac{1}{2} \times \eta \times e^{(t)} \times \overline{o}^{j,(t)}\right) \qquad (۶۲)$$

$$s^{j,(t+1)} = s^{j,(t)} + \Delta s^{j,(t)} = s^{j,(t)} + \eta e^{(t)} o^{j,(t)} \qquad (۶۳)$$

در لایهٔ میانی سه پارامتر وابسته به تابع فعال‌ساز گرانولی آموزش می‌بینند که الف) مرکز دسته، ب) کران پائینِ انحراف استاندارد و ج) کران بالای انحراف استاندارد می‌باشند.

برای آموزش مرکزِ توابع گاوسی که از پارامترهای لایهٔ میانی است و با فرض $\overline{o}^{j,(t)} \ge \underline{o}^{j,(t)}$ و $\overline{d}^{j,(t)} \ge \underline{d}^{j,(t)}$ خواهیم داشت[۴، ۵، ۳۸]:

$$\Delta c_r^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial c_r^{j,(t)}} = \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \underline{y}^{(t)}} \times \frac{\partial \underline{y}^{(t)}}{\partial \underline{o}^{j,(t)}} \times \frac{\partial \underline{o}^{j,(t)}}{\partial \overline{d}^{j,(t)}} \times \frac{\partial \overline{d}^{j,(t)}}{\partial c_r^{j,(t)}}\right) + \left(-\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \overline{y}^{(t)}} \times \frac{\partial \overline{y}^{(t)}}{\partial \overline{o}^{j,(t)}} \times \frac{\partial \overline{o}^{j,(t)}}{\partial \underline{d}^{j,(t)}} \times \frac{\partial \underline{d}^{j,(t)}}{\partial c_r^{j,(t)}}\right) \qquad (۶۴)$$

$$\Delta c_r^{j,(t)} = \left[-\eta \times e^{(t)} \times \left(-\frac{1}{2}\right) \times (v^{j,(t)} - s^{j,(t)}) \times \left(-\frac{1}{2} \times \underline{o}^{j,(t)}\right) \times \left(-2 \times \frac{\left(\overline{u}_r^{j,(t)}\right)^2}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)}\right)\right] + \left[-\eta \times e^{(t)} \times \left(-\frac{1}{2}\right) \times (v^{j,(t)} + s^{j,(t)}) \times \left(-\frac{1}{2} \times \overline{o}^{j,(t)}\right) \times \left(-2 \times \frac{\left(\underline{u}_r^{j,(t)}\right)^2}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)}\right)\right] \qquad (۶۵)$$

$$\Delta c_r^{j,(t)} = \left[\frac{1}{2} \times \eta \times e^{(t)} \times \frac{\underline{v}^{j,(t)} \times \underline{o}^{j,(t)} \times \left(\overline{u}_r^{j,(t)}\right)^2}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)}\right] + \left[\frac{1}{2} \times \eta \times e^{(t)} \times \frac{\overline{v}^{j,(t)} \times \overline{o}^{j,(t)} \times \left(\underline{u}_r^{j,(t)}\right)^2}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)}\right] \qquad (۶۶)$$

$$\Delta c_r^{j,(t)} = \left[\frac{1}{2} \times \eta \times e^{(t)} \times \frac{\left[\underline{v}^{j,(t)} \times \underline{o}^{j,(t)} \times \left(\overline{u}_r^{j,(t)}\right)^2\right] + \left[\overline{v}^{j,(t)} \times \overline{o}^{j,(t)} \times \left(\underline{u}_r^{j,(t)}\right)^2\right]}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)}\right] \qquad (۶۷)$$

$$c_r^{j,(t+1)} = c_r^{j,(t)} + \Delta c_r^{j,(t)} = c_r^{j,(t)} + \frac{1}{2}\eta e^{(t)} \frac{\left[\underline{v}^{j,(t)} \times \underline{o}^{j,(t)} \times \left(\overline{u}_r^{j,(t)}\right)^2\right] + \left[\overline{v}^{j,(t)} \times \overline{o}^{j,(t)} \times \left(\underline{u}_r^{j,(t)}\right)^2\right]}{\left(x_r^{(t)} + n_r^{(t)} - c_r^{j,(t)}\right)} \qquad (۶۸)$$

برای آموزش کرانِ پائینِ انحراف استاندارد که از پارامترهای لایهٔ میانی است با فرض $\overline{o}^{j,(t)} \ge \underline{o}^{j,(t)}$ و $\overline{d}^{j,(t)} \ge \underline{d}^{j,(t)}$، خواهیم داشت:

$$\Delta \underline{\sigma}_r^{j,(t)} = -\eta \times \frac{\partial E^{(t)}}{\partial \underline{\sigma}_r^{j,(t)}} = -\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \underline{y}^{(t)}} \times \frac{\partial \underline{y}^{(t)}}{\partial \underline{o}^{j,(t)}} \times \frac{\partial \underline{o}^{j,(t)}}{\partial \overline{d}^{j,(t)}} \times \frac{\partial \overline{d}^{j,(t)}}{\partial \underline{\sigma}_r^{j,(t)}} \qquad (۶۹)$$

$$\Delta \underline{\sigma}_r^{j,(t)} = -\eta \times e^{(t)} \times \left(-\tfrac{1}{2}\right) \times (v^{j,(t)} - s^{j,(t)}) \times \left(-\tfrac{1}{2} \times \underline{o}^{j,(t)}\right) \times \left[-2 \times \frac{\left(\overline{u}_r^{j,(t)}\right)^2}{\underline{\sigma}_r^{j,(t)}}\right] = \frac{1}{2} \times \eta \times e^{(t)} \times \underline{v}^{j,(t)} \times \frac{\underline{o}^{j,(t)} \times \left(\overline{u}_r^{j,(t)}\right)^2}{\underline{\sigma}_r^{j,(t)}} \qquad (۷۰)$$

$$\underline{\sigma}_r^{j,(t+1)} = \underline{\sigma}_r^{j,(t)} + \Delta \underline{\sigma}_r^{j,(t)} = \underline{\sigma}_r^{j,(t)} + \frac{1}{2}\eta e^{(t)} \underline{v}^{j,(t)} \frac{\underline{o}^{j,(t)} \times \left(\overline{u}_r^{j,(t)}\right)^2}{\underline{\sigma}_r^{j,(t)}} \qquad (۷۱)$$

برای آموزش کرانِ بالای انحراف استاندارد که از پارامترهای لایهٔ میانی است با فرض $\overline{o}^{j,(t)} \ge \underline{o}^{j,(t)}$ و $\overline{d}^{j,(t)} \ge \underline{d}^{j,(t)}$، خواهیم داشت:

$$\Delta \overline{\sigma}_r^{j,(t)} = -\eta \frac{\partial E^{(t)}}{\partial \overline{\sigma}_r^{j,(t)}} = -\eta \times \frac{\partial E^{(t)}}{\partial e^{(t)}} \times \frac{\partial e^{(t)}}{\partial \overline{y}^{(t)}} \times \frac{\partial \overline{y}^{(t)}}{\partial \overline{o}^{j,(t)}} \times \frac{\partial \overline{o}^{j,(t)}}{\partial \underline{d}^{j,(t)}} \times \frac{\partial \underline{d}^{j,(t)}}{\partial \overline{\sigma}_r^{j,(t)}} \qquad (۷۲)$$

$$\Delta\overline{\sigma}_r^{j,(t)} = -\eta \times e^{(t)} \times \left(-\tfrac{1}{2}\right) \times \left(v^{j,(t)} + s^{j,(t)}\right) \times \left(-\tfrac{1}{2} \times \overline{\sigma}^{j,(t)}\right) \times \left(-2 \times \frac{\left(\underline{u}_r^{j,(t)}\right)^2}{\overline{\sigma}_r^{j,(t)}}\right) = \tfrac{1}{2} \times \eta \times e^{(t)} \times \overline{v}^{j} \times \frac{\overline{\sigma}^{j,(t)} \times \left(\underline{u}_r^{j,(t)}\right)^2}{\overline{\sigma}_r^{j,(t)}} \quad (۷۳)$$

$$\overline{\sigma}_r^{j,(t+1)} = \overline{\sigma}_r^{j,(t)} + \Delta\overline{\sigma}_r^{j,(t)} = \overline{\sigma}_r^{j,(t)} + \frac{1}{2}\eta e^{(t)} \overline{v}^{j,(t)} \frac{\overline{\sigma}^{j,(t)} \times \left(\underline{u}_r^{j,(t)}\right)^2}{\overline{\sigma}_r^{j,(t)}} \quad (۷۴)$$

۵- نتایج شبیه‌سازی

۵-۱ شناسایی سیستمِ غیر خطیِ پویایِ U شکل با پنج ورودی

هدف شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» است که توسط معادلات زیر تولید می‌گردد[۶].

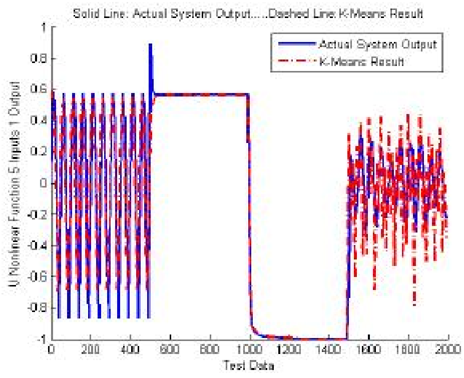
$$y_p(k+1) = f\left(y_p(k), y_p(k-1), y_p(k-2), u(k), u(k-1)\right) \quad (۷۵)$$

که در آن:

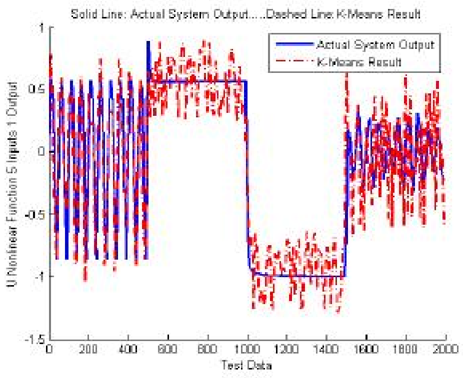
$$f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_2^2 + x_3^2} \quad (۷۶)$$

$$u(k) = \begin{cases} \sin\left(\frac{\pi k}{25}\right) & 0 < k < 500 \\ +1 & 500 \le k < 1000 \\ -1 & 1000 \le k < 1500 \\ 0.3 \sin\left(\frac{\pi k}{25}\right) + 0.1 \sin\left(\frac{\pi k}{32}\right) + 0.6 \sin\left(\frac{\pi k}{10}\right) & 1500 \le k \le 2000 \end{cases} \quad (۷۷)$$

در شکل۵، رنگ «آبی و ممتد» نشان دهندهٔ خروجی واقعی سیستم و رنگ «قرمز و خط چین» نشان دهندهٔ شناسایی سیستم به کمک روش «الگوریتم خوشه‌بندی K-Means» است.



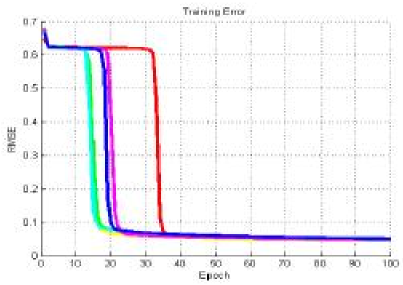
(الف)



(ب)

شکل۵. شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» توسط شبکهٔ عصبیِ RBF گرانولی با ۲۰ نُرون لایهٔ میانی و الف) بدون نویز ب) $SNR=0$

در شکل۶ خطای آموزش شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» برای ۶ بار اجرای برنامه آورده شده است.



شکل۶. خطای آموزش شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» توسط شبکهٔ عصبیِ RBF گرانولی با ۲۰ نُرون لایهٔ میانی و $SNR=0$

تعداد داده‌های به‌کار رفته برای آموزش در این مثال ۱۵۰۰ داده است. در این مسأله ۴۴۲ داده به عنوان دادهٔ آزمایشی به کار رفته‌اند.

در جدول۱ نتایج شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» توسط شبکهٔ عصبیِ RBF گرانولی و شبکهٔ عصبیِ RBF برای ۱۰۰ بار تکرار الگوریتم آورده شده است.

جدول ۱. نتایج شناسایی «سیستمِ غیر خطیِ پویایِ U شکل با پنج ورودی» توسط شبکهٔ عصبیِ RBF گرانولی و شبکهٔ عصبیِ RBF

SNR (دسی بل)	خطا (RMS)	شبکهٔ عصبی RBF گرانولی					شبکهٔ عصبی RBF	
		گرادیان نزولی			K-Means		گرادیان نزولی	
		تعداد نُرون‌های لایهٔ پنهان			تعداد نُرون‌های لایهٔ		تعداد نُرون‌های لایهٔ	
		۱	۱۰	۲۰	۱۰	۲۰	۱۰	۲۰
۰	آموزش	۰/۳۷۸۳	۰/۰۶۷۲	۰/۰۶۴۹	۰/۰۷۱۲	۰/۰۷۰۰	۰/۱۸۷۶	۰/۱۷۷۵
	آزمون	۰/۵۱۳۴	۰/۴۳۳۴	۰/۴۳۱۰	۰/۴۳۷۶	۰/۴۳۶۵	۰/۶۳۹۲	۰/۶۰۴۵
۵	آموزش	۰/۲۷۸۱	۰/۰۶۱۹	۰/۰۵۸۰	۰/۰۶۰۵	۰/۰۵۵۸	۰/۱۵۸۴	۰/۱۵۵۵
	آزمون	۰/۳۸۴۶	۰/۳۵۵۳	۰/۳۵۲۶	۰/۳۵۴۲	۰/۳۵۱۵	۰/۴۴۴۱	۰/۴۴۰۰
۱۰	آموزش	۰/۲۳۵۰	۰/۰۵۷۵	۰/۰۵۳۳	۰/۰۵۵۵	۰/۰۵۰۰	۰/۱۰۵۰	۰/۰۹۷۵
	آزمون	۰/۳۳۹۴	۰/۲۰۶۹	۰/۲۰۴۱	۰/۲۰۵۴	۰/۲۰۳۰	۰/۳۰۲۸	۰/۲۸۰۰
بدون نویز	آموزش	۰/۲۲۲۱	۰/۰۴۹۲	۰/۰۴۶۷	۰/۰۴۳۲	۰/۰۴۰۰	۰/۰۴۷۰	۰/۰۴۴۵
	آزمون	۰/۳۰۵۶	۰/۰۵۷۲	۰/۰۵۳۲	۰/۰۵۵۱	۰/۰۵۰۵	۰/۰۵۶۰	۰/۰۵۲۱

همانطور که از نتایج مندرج در جدول۱ معلوم می گردد، در شرایطی که داده‌های ورودی آغشته به نویز نیستند، عملکرد شبکهٔ عصبیِ RBF

گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» نسبت به عملکرد همین شبکه با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» و شبکهٔ عصبیِ RBF با آموزش تمام پارامترها بوسیلهٔ روش «گرادیان نزولی»، تا حدّی بهتر می‌باشد. دلیل این مسأله این است که در شناسایی سیستم‌های غیر خطیِ پویا و در شرایطی که داده‌های ورودی آغشته به نویز نیستند؛ تعیین تعلق یک دسته دادهٔ ورودی به یک خوشه به درستی و به سرعت صورت می‌پذیرد. و این تعیین دقیق خوشه در لایهٔ پنهان، مبنای شناساییِ درست و سریع در لایهٔ خروجی قرار خواهد گرفت. در حالیکه الگوریتم «گرادیان نزولی» به دلیل احتمال گیر افتادن در مینیمم محلّی و دشواری یافتن نرخ آموزش مناسب، در تعداد دفعاتِ تکرارِ مساوی، در شناساییِ سیستم‌های غیرِ خطیِ پویا ضعیفتر عمل می‌کند.

در شرایط بدون نویز، برتری جزیی عملکرد شبکهٔ عصبیِ RBF نسبت به شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی»، به دلیل تقریب ناشی از محاسبات بازه‌ای است که روش «گرادیان نزولی» در شبکهٔ عصبیِ RBF گرانولی را با خطایِ حاصل از تقریب روبرو می‌سازد.

پس در شرایط بدون نویز، شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» بهترین عملکرد را دارد و پس از آن شبکهٔ عصبیِ RBF و در نهایت شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» قرار می‌گیرد.

در شرایطی که داده‌های ورودی آغشته به نویز ضعیف و متوسط هستند نیز برتری روش «الگوریتم خوشه‌بندی K-Means» البته به مقدار کمتر[۱] ادامه می‌یابد. برتری کمتر «الگوریتم خوشه‌بندی K-Means» در حضور نویز به این دلیل است که هر چه دامنهٔ نویز بیشتر می‌گردد؛ تعیین تعلق دسته دادهٔ ورودی به یک خوشه مشکلتر می‌شود. ولی خطای بوجود آمده از این مسأله هنوز آن اندازه بزرگ نشده که بر نقاط ضعف الگوریتم «گرادیان نزولی» غلبه نماید.

هم‌چنین مشاهده می‌گردد که هر چه میزان نویز تزریق شده به داده‌ها بیشتر باشد، عملکرد شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» نسبت به شبکهٔ عصبیِ RBF بهتر می‌شود. علت این امر این است که خطایِ حاصل از تقریب در محاسبات بازه‌ای آنقدر بزرگ نیست که سردرگمی شبکهٔ عصبیِ RBF در مقابل داده‌های ورودی آغشته به نویز را جبران کند. در نتیجه خطایِ شناساییِ شبکهٔ عصبیِ RBF، با افزایش دامنهٔ نویز به سرعت بزرگتر می‌گردد و این شبکهٔ عصبی توانایی شناسایی خود را هر چه بیشتر از دست می‌دهد.

پس در شرایطی که داده‌های ورودی آغشته به نویز ضعیف و متوسط هستند، نیز شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» هم چنان بهترین عملکرد را دارد ولی پس از آن شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» و دست آخر شبکهٔ عصبیِ RBF قرار می‌گیرد.

نتایج موجود در جدول۱ همچنین نشان می‌دهد که در نویز با «نسبت سیگنال به نویز» برابر صفر، عملکرد شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» بهتر از حالتی است که پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» آموزش می‌بینند. چرا که در هنگام وجود نویز قوی در دادهٔ ورودی، تعیین تعلق آن به یک خوشه دشوار می‌باشد و خطای بوجود آمده در تعیین خوشه بر تقریب ناشی از محاسبات بازه‌ای غلبه می‌یابد.

پس در وضعیت وجود نویز قوی در دادهٔ ورودی؛ شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» بهترین عملکرد را دارد و پس از آن شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» و دست آخر شبکهٔ عصبیِ RBF قرار می‌گیرد.

جدول۲ مقایسهٔ نتایج شناسایی «سیستم غیر خطیِ پویایِ U شکل با پنج ورودی» بوسیلهٔ شبکهٔ عصبیِ RBF گرانولی، با شبکهٔ عصبی فازی که آموزش پارامترهای مقدّم و تالی در آن به دو روش «بهینه‌سازی گروهی ذرّات عمومی[۲]» و «بهینه‌سازی گروهی ذرّات بهبود یافته[۳]» صورت پذیرفته را نشان می‌دهد.[۳۶]

جدول۲. نتایج شناساییِ «سیستمِ غیر خطیِ پویایِ U شکل با پنج ورودی» توسط شبکهٔ عصبیِ RBF گرانولی و شبکه‌هایِ عصبی- فازی دیگر

SNR (دسی بل)	خطا (RMS)	شبکهٔ عصبیِ RBF گرانولی (۳۰ نُرون لایهٔ میانی) 100 epochs		شبکهٔ عصبیِ RBF گرانولی (۴۰ نُرون لایهٔ میانی) 100 epochs		شبکهٔ عصبی- فازی با آموزش پارامترهای مقدّم و تالی به روش «بهینه سازی گروهی ذرّات عمومی» 100 epochs	شبکهٔ عصبی- فازی با آموزش پارامترهای مقدّم و تالی به روش «بهینه سازی گروهی ذرّات بهبود یافته» 100 epochs
		گرادیان نزولی	K-Means	گرادیان نزولی	K-Means		
بدون نویز	آموزش	۰/۰۴۴۹	۰/۰۳۹۱	۰/۰۴۱۴	۰/۰۳۷۶	۰/۱۹۷	۰/۰۰۰۷۰
بدون نویز	آزمون	۰/۰۵۰۹	۰/۰۴۹۷	۰/۰۴۸۲	۰/۰۴۷۹	؟	؟

[۱] این مقدار وابسته به دامنهٔ نویز است

[2] General Particular Swarm Organization (General PSO)
[3] Modified Particular Swarm Organization (Modified PSO)

خطای آزمون برای این شبکهٔ عصبی فازی در دسترس نبود. ولی از مقایسهٔ خطای آموزش می‌توان پی برد که روند شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» بوسیلهٔ شبکهٔ عصبی RBF گرانولی فارغ از نوع آموزش پارامترهای آن؛ نسبت به شبکهٔ عصبی فازی با آموزش پارامترهای مقدّم و تالی به روش «بهینه‌سازی گروهی ذرّات عمومی» بهتر صورت می‌پذیرد.

هر چند که در ۱۰۰ بار تکرار الگوریتم، شبکهٔ عصبی فازی با آموزش پارامترهای مقدّم و تالی به روش «بهینه‌سازی گروهی ذرّات بهبود یافته» توانایی بیشتری را در شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» نشان می‌دهد. ولی به علت زیاد بودن تعداد پارامترها و کندی الگوریتم‌های تکاملی از جمله الگوریتم تکاملی «بهینه‌سازی گروهی ذرّات بهبود یافته»، به نظر می‌رسد که این الگوریتم‌ها باید زمان اجرای بزرگتری داشته باشند.

۵-۲ پیش‌بینی سری زمانی مکی‌گلاس با توجه به وجود نویز ورودی

این سری زمانی توسط معادلهٔ زیر تولید می‌گردد[۳۰].

$$\dot{x} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \qquad (۷۸)$$

از ۱۲۰۱ داده تولید شده به وسیلهٔ این سری زمانی؛ تعداد ۸۴۰ داده به عنوان دادهٔ آموزشی[1] و ۳۵۲ تا به عنوان دادهٔ آزمایشی[2] به کار رفته‌اند.

جدول۳. نتایج پیش‌بینی سری زمانی مکی‌گلاس توسط شبکهٔ عصبی RBF گرانولی و شبکهٔ عصبی RBF

SNR (دسی بل)	خطا (RMS)	شبکهٔ عصبی RBF گرانولی					شبکهٔ عصبی RBF	
		گرادیان نزولی			K-Means		گرادیان نزولی	
		تعداد نُرون‌های لایهٔ پنهان			تعداد نُرون‌های لایهٔ		تعداد نُرون‌های لایهٔ	
		۱	۱۰	۲۰	۱۰	۲۰	۱۰	۲۰
۰	آموزش	۰/۰۵۳۲	۰/۰۲۲۸	۰/۰۲۰۷	۰/۰۴۵۲	۰/۰۴۰۰	۰/۱۱۳۲	۰/۱۰۸۶
	آزمون	۰/۲۱۱۱	۰/۲۱۰۹	۰/۲۱۰۷	۰/۳۴۰۸	۰/۳۲۳۶	۰/۹۲۲۵	۰/۸۴۳۶
۵	آموزش	۰/۰۳۸۲	۰/۰۲۰۶	۰/۰۱۵۷	۰/۰۳۴۵	۰/۰۳۱۵	۰/۰۵۶۶	۰/۰۵۴۵
	آزمون	۰/۱۱۸۳	۰/۱۱۹۰	۰/۱۱۷۰	۰/۲۴۰۰	۰/۲۲۷۵	۰/۴۴۷۴	۰/۳۰۵۷
۱۰	آموزش	۰/۰۳۴۵	۰/۰۱۸۳	۰/۰۱۳۸	۰/۰۲۹۳	۰/۰۲۷۵	۰/۰۳۱۷	۰/۰۳۰۴
	آزمون	۰/۰۶۹۱	۰/۰۶۶۸	۰/۰۶۶۲	۰/۰۹۹۵	۰/۰۷۸۰	۰/۱۱۲۸	۰/۱۰۰۲
بدون نویز	آموزش	۰/۰۳۱۰	۰/۰۱۱۳	۰/۰۱۱۱	۰/۰۱۱۱	۰/۰۱۱۰	۰/۰۱۱۳	۰/۰۱۱۰
	آزمون	۰/۰۳۲۴	۰/۰۱۳۷	۰/۰۱۲۸	۰/۰۱۳۱	۰/۰۱۲۷	۰/۰۱۳۷	۰/۰۱۲۷

همانطور که از نتایج مندرج در جدول۳ معلوم می‌گردد؛ در شرایطی که داده‌های ورودی آغشته به نویز نیستند، عملکرد شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه-بندی K-Means» نسبت به عملکرد همین شبکه با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» و شبکهٔ عصبی RBF با آموزش تمام پارامترها بوسیلهٔ روش «گرادیان نزولی»، اندکی بهتر می‌باشد.

دلیل این مسأله این است که در پیش‌بینی توابع آشوب در افق زمانی بزرگ و در شرایطی که داده‌های ورودی آغشته به نویز نیستند؛ تعیین تعلق یک دسته دادهٔ ورودی به یک خوشه در لایهٔ پنهان درست و سریع صورت می‌پذیرد. حال از آن جایی که در بحث پیش‌بینی فرض بر آن است که دو بردار ورودی که فاصلهٔ اقلیدسی کمی دارند؛ خروجی‌های نزدیک به هم تولید کنند تا سیستم پیش‌بینی‌پذیر گردد. ولی در توابع آشوب در افق زمانی بزرگ، دو بردار ورودی که فاصلهٔ اقلیدسی کمی دارند؛ ممکن است که خروجی‌های دور از هم تولید کنند. پس باید پارامترهایی همانند وزن‌های بازه‌ای لایهٔ خروجی؛ درجهٔ آزادی شبکهٔ عصبی RBF گرانولی را به اندازه‌ای بالا ببرند که به کمک آموزش وزن-های بازه‌ای لایهٔ خروجی؛ بتوان این ویژگی توابع آشوب را پوشش داد و تولید خروجی‌های دور از هم را پیش‌بینی نمود. در نهایت در نبود نقاط ضعف روش «گرادیان نزولی» یعنی احتمال گیر افتادن در مینیمم محلّی و دشواری یافتن نرخ آموزش مناسب؛ برتری در شرایط فاقد نویز با روش «الگوریتم خوشه‌بندی K-Means» خواهد بود.

سرانجام در شبکهٔ عصبی RBF، به علت بازه‌ای نبودن وزن‌های لایهٔ خروجی؛ انعطاف‌پذیری شبکهٔ عصبی برای پیش‌بینی تولید خروجی‌های دور از هم در توابع آشوب؛ کمتر شده و همین مسأله وجود خطای ناشی از تقریب در محاسبات بازه‌ای در شبکهٔ عصبی RBF گرانولی را پوشش داده و دو شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» و شبکهٔ عصبی RBF با آموزش تمام پارامترها بوسیلهٔ روش «گرادیان نزولی» را به نتایج مشابهی می‌رساند. در حالی که شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» به علت رهایی از روش «گرادیان نزولی» در آموزش پارامترهای لایهٔ پنهان؛ به نتایج بهتری دست پیدا می‌کند.

پس در شرایط بدون نویز، شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» بهترین عملکرد را دارد و پس از آن شبکهٔ عصبی RBF و شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» در یک رده قرار می‌گیرند.

علت برتری روش «گرادیان نزولی» در شبکهٔ عصبی RBF گرانولی؛ و در هنگام شناسایی و پیش‌بینی توابع آشوب در افق زمانی بزرگ؛ به ویژه در حضور نویز با «نسبت سیگنال به نویز» برابر صفر و پنج و تا حدّ کمتری ده، این است که چون در آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means»، و با حضور نویز متوسط و قوی و تا حدّی نویز ضعیف، تعیین تعلق دسته دادهٔ ورودی به یک خوشه به درستی و به سرعت صورت نگرفته است؛ پس آموزش وزن‌های بازه‌ای لایهٔ خروجی در شناسایی و پیش‌بینی درست توابع آشوب کارساز نیست

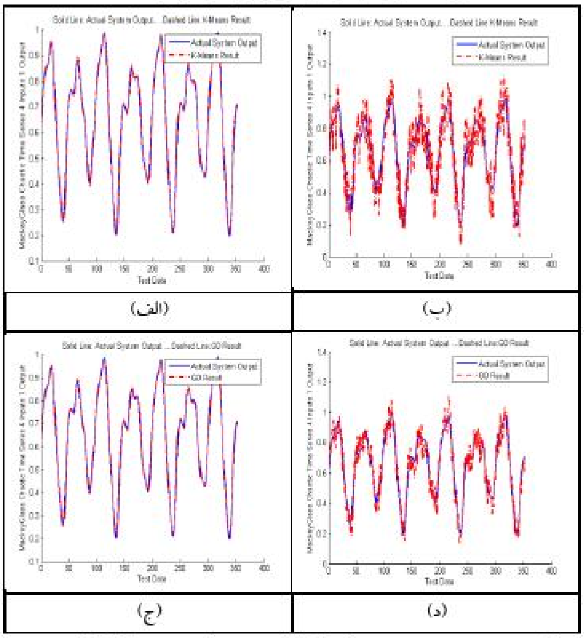
[1] Training data set
[2] Test data set

و زمان کافی برای آموزش این وزن‌ها در تعداد تکرار محدود الگوریتم وجود ندارد. از اینرو می‌توان گفت که خطای بوجود آمده در تعیین خوشه در «الگوریتم خوشه‌بندی K-Means»، آنقدر بزرگ است که بر خطای تقریب ناشی از محاسبات بازه‌ای، در «الگوریتم گرادیان نزولی» برتری می‌یابد و روش «گرادیان نزولی» بهتر عمل می‌نماید.

سرانجام در مورد شبکهٔ عصبی RBF می‌توان گفت که در مقابل داده‌های ورودی آغشته به نویز، دچار سردرگمی در شناسایی و پیش‌بینی توابع آشوب می‌گردد. زیرا به علت گرانولی نبودن توابع فعّال‌ساز در لایهٔ پنهان و بازه‌ای نبودن وزن‌های لایهٔ خروجی، انعطاف پذیری کافی در مواجهه با نویز را ندارد.

پس در وضعیت وجود نویز در دادهٔ ورودی، شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «گرادیان نزولی» بهترین عملکرد را دارد و پس از آن شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» و دست آخر شبکهٔ عصبی RBF قرار می‌گیرد.

در شکل۷، رنگ «آبی و ممتد» نشان دهندهٔ خروجی واقعی سیستم و رنگ «قرمز و خط چین» نشان دهندهٔ پیش‌بینی سری زمانی مکی‌گلاس توسط شبکهٔ عصبی RBF گرانولی به کمک «الگوریتم خوشه‌بندی K-Means» و «الگوریتم گرادیان نزولی» است. همانطور که انتظار می‌رود در وضعیت عدم وجود نویز در دادهٔ ورودی، «الگوریتم خوشه‌بندی K-Means» عملکرد بهتری دارد در حالی که در حضور نویز قوی در دادهٔ ورودی، عملکرد «الگوریتم گرادیان نزولی» بهتر می‌گردد.



شکل۷. پیش‌بینی سری زمانی مکی‌گلاس توسط شبکهٔ عصبی RBF گرانولی با ۲۰ نُرون لایهٔ میانی و به روش الف) «الگوریتم خوشه‌بندی K-Means» و بدون نویز ب) «الگوریتم خوشه‌بندی K-Means» و SNR=0 ج) «الگوریتم گرادیان نزولی» و بدون نویز د) «الگوریتم گرادیان نزولی» و SNR=0

در جدول۴ مقایسهٔ خطای آزمون پیش‌بینی «سری زمانی مکی-گلاس» توسط «شبکهٔ عصبی RBF گرانولی» با «سیستم‌های منطق فازی نوع۱ و نوع۲ [۱۱]»آورده شده است.

جدول۴. نتایج پیش‌بینی سری زمانی مکی‌گلاس توسط شبکهٔ عصبی RBF گرانولی و سیستم‌های منطق فازی نوع۱ و نوع۲

SNR (دسی بل)	خطا (RMS)	شبکهٔ عصبی RBF گرانولی (بیست نُرون لایهٔ میانی)		شبکهٔ عصبی RBF گرانولی (دویست نُرون لایهٔ میانی)		سیستم‌های منطق فازی		
		گرادیان نزولی	K-Means	گرادیان نزولی	K-Means	نوع۲		نوع۱
						تابع تعلق گاوسی	تابع تعلق الماس شکل	
۰	آزمون	۰/۲۱۰۷	۰/۳۳۳۶	۰/۱۱۰۵	۰/۲۰۱۵	۰/۱۴۲۹	۰/۱۴۲۲	۰/۱۵۱۷
۱۰	آزمون	۰/۰۶۶۲	۰/۰۷۸۰	۰/۰۱۲۱	۰/۰۱۱۸	۰/۰۸۲۸	۰/۰۹۳۵	۰/۰۹۵۳

همان طور که از مقایسهٔ نتایج مندرج در جدول۴ مشخص است؛ در حضور نویز ضعیف عملکرد «شبکهٔ عصبی RBF گرانولی» با هر دو روش آموزش؛ بهتر از «سیستم‌های منطق فازی نوع۱ و نوع۲» می‌باشد. که با توجه به بیشتر بودن تعداد پارامترهای آموزش دیده در «سیستم‌های منطق فازی نوع۱ و بویژه نوع۲» نسبت به «شبکهٔ عصبی RBF گرانولی»، و پیچیدگی بیشتر الگوریتم آموزشی آن‌ها، زمان اجرای برنامه در «سیستم‌های منطق فازی نوع۱ و نوع۲» به شدّت افزایش می‌یابد. و این موضوع برتری «شبکهٔ عصبی RBF گرانولی» را برجسته‌تر می‌سازد.

در حضور نویز قوی، عملکرد «شبکهٔ عصبی RBF گرانولی» با افزایش تعداد نُرون‌های لایهٔ میانی؛ می‌تواند نسبت به عملکرد «سیستم‌های منطق فازی نوع۱ و نوع۲» برتری یابد. زیرا با افزایش تعداد نُرون‌های لایهٔ میانی؛ درجهٔ آزادی سیستم با افزایش تعداد پارامترهای آموزش دیده، زیاد می‌گردد. به عنوان مثال با داشتن ۳۰۰ نُرون در لایهٔ پنهان، خطای آزمون برای روش «الگوریتم خوشه‌بندی K-Means»، به ۰/۱۵ کاهش می‌یابد که از «سیستم‌های منطق فازی نوع۱» بهتر است. البته زمان اجرای برنامه به مقدار زیادی افزایش می‌یابد ولی در مقایسه با با زمان اجرای چند دقیقه‌ای «سیستم‌های منطق فازی نوع۲»، زمان اجرای کمتر از یک دقیقه(جدول۷) در شبکهٔ عصبی RBF گرانولی؛ مناسب به نظر می‌رسد.

در جدول۵ مقایسهٔ خطای آموزشِ پیش‌بینیِ «سریِ زمانیِ مکی-گلاس»، توسط «شبکهٔ عصبیِ RBF گرانولی» با «شبکهٔ عصبیِ RBF قوی با هرس و رشد نُرون‌های لایهٔ میانی [۱۶]»، آورده شده است.

جدول۵. نتایج پیش‌بینیِ سریِ زمانیِ مکی‌گلاس توسط شبکهٔ عصبیِ RBF گرانولی و شبکهٔ عصبیِ RBF قوی با هرس و رشد نُرون‌های لایهٔ میانی

SNR (دسی بل) در قیاس با (درصد سطح نویز)	خطا (RMS)	شبکهٔ عصبیِ RBF گرانولی (ده نرون لایهٔ میانی) 100 epochs		شبکهٔ عصبیِ RBF گرانولی (بیست نرون لایهٔ میانی) 100 epochs		شبکهٔ عصبیِ RBF قوی با هرس و رشد نُرون‌های لایهٔ میانی 500 epochs (۸ نرون لایهٔ میانی در مرحلهٔ آخر)	شبکهٔ عصبیِ RBF 100 epochs	
		گرادیان نزولی	K-Means	گرادیان نزولی	K-Means		ده نُرون لایهٔ میانی	بیست نُرون لایهٔ میانی
۵ > ٪۴۰	آموزش	۰/۰۲۰۶	۰/۰۳۴۵	۰/۰۱۵۷	۰/۰۳۱۵	/۰۴۷۲۸۵	۰/۰۵۶۶	۰/۰۵۴۵
۱۰≈ ٪۲۰	آموزش	۰/۰۱۸۳	۰/۰۲۹۳	۰/۰۱۳۸	۰/۰۲۷۵	/۰۲۷۰۳۵	۰/۰۳۱۷	۰/۰۳۰۴

خطایِ آزمون برای این شبکهٔ عصبی موجود نبود ولی از مقایسهٔ خطایِ آموزش می‌توان پی برد که روند پیش‌بینیِ «سریِ زمانیِ مکی-گلاس»، بوسیلهٔ شبکهٔ عصبیِ RBF گرانولی فارغ از نوع آموزش پارامترهایِ آن، نسبت به «شبکهٔ عصبیِ RBF قوی با هرس و رشد نُرون-های لایهٔ میانی»، بهتر صورت می‌پذیرد. و این برتری با قوی‌تر شدن نویز، نمود بیشتری می‌یابد. علت این امر وجود تابع فعال‌ساز گرانولی در ساختار شبکهٔ عصبیِ RBF گرانولی است که با تنظیم «پارامترهای مربوط به پهنایِ عدم قطعیّت، یعنی کران پائین و کران بالای انحراف استاندارد»، مدیریت عدم قطعیت در شبکهٔ عصبیِ RBF گرانولی را انجام می‌دهد.

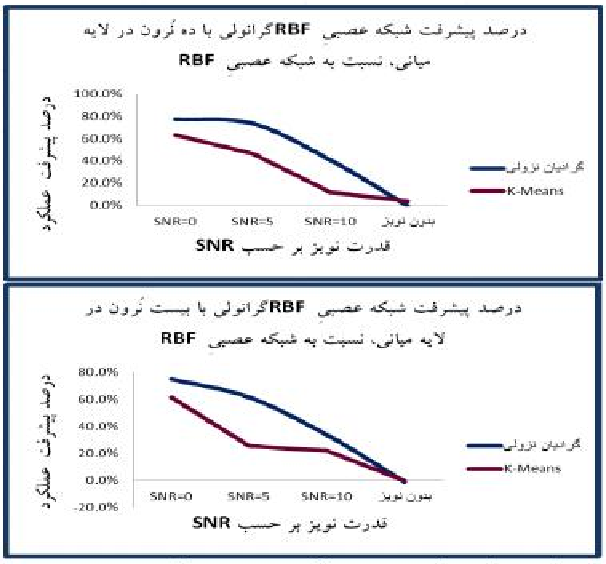
در جدول۶ درصد پیشرفت عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF در هنگام کاهش میزان نویز تزریق شده به داده‌ها آورده شده است. در شبکهٔ عصبیِ RBF گرانولی، هر دو روش «گرادیان نزولی» و «الگوریتم خوشه‌بندی K-Means»، برای آموزش پارامترهای لایهٔ میانی مد نظر قرار گرفته است. خطای آزمون مربوط به پیش‌بینیِ «سریِ زمانیِ آشوبِ مکی گلاس» می‌باشد و در توابع آشوب دیگر مانند «شناساییِ جاذبِ لورنز[۱] [۳۲]» و «شناساییِ نگاشتِ آشوبگونه ایکدا[۲]» و «شناساییِ نگاشتِ هنون[۳] [۳۱]» ، نیز انتظار این است که همین رفتار مشاهده گردد[۲۷ ۳۲].

جدول۶. درصد بهبود عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF در پیش‌بینیِ سریِ زمانیِ آشوبِ مکی گلاس

SNR (دسی بل)	خطای آزمون (بیست نُرون لایهٔ میانی)			خطای آزمون (ده نُرون لایهٔ میانی)		
	شبکهٔ عصبیِ گرانولی RBF		شبکهٔ عصبیِ RBF	شبکهٔ عصبیِ گرانولی RBF		شبکهٔ عصبیِ RBF
	گرادیان نزولی	K-Means	گرادیان نزولی	گرادیان نزولی	K-Means	گرادیان نزولی
۰	/۲۱۰۷	۰/۳۲۳۶	/۸۴۳۶	۲۱۰۹	۰/۳۴۰۸	۰/۹۲۲۵
۵	/۱۱۷۰	۰/۲۲۷۵	/۳۰۵۷	۱۱۹۰	۰/۲۴۰۰	۰/۴۴۷۴
۱۰	/۰۶۶۲	۰/۰۷۸۰	/۱۰۰۲	/۰۶۶۸	۰/۰۹۹۵	۰/۱۱۲۸
بدون نویز	/۰۱۲۸	۰/۰۱۲۷	/۰۱۲۷	/۰۱۳۷	۰/۰۱۳۱	۰/۰۱۳۷

SNR (دسی بل)	درصد بهبود عملکرد		درصد بهبود	
	گرادیان نزولی	K-Means	گرادیان نزولی	K-Means
۰	٪۷۵/۰	٪۶۱/۶	۷۷/۱	٪۶۳/۱
۵	٪۶۱/۷	٪۲۵/۶	۷۳/۴	٪۴۶/۴
۱۰	٪۳۳/۹	٪۲۲/۲	۴۰/۸	٪۱۱/۸
بدون نویز	-٪۰/۸	٪۰	٪۰	٪۴/۴

در شکل۸ که مربوط به پیش‌بینیِ «سریِ زمانیِ آشوبِ مکی گلاس» می‌باشد، نمودار کاهش درصد پیشرفت عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF در هنگام کاهش میزان نویز تزریق شده به داده‌های ورودی و برای هر دو روش آموزشی، نشان داده شده است.



شکل۸. عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF

[1] Lorenz attractor
[2] Daisaku Ikeda
[3] Henon map

۵-۳ عوامل مؤثر بر زمان اجرای[1] برنامه

همان‌طور که از نتایج به دست آمده از شناسایی «سیستم غیر خطی پویای U شکل با پنج ورودی» و پیش‌بینی «سری زمانی مکی‌گلاس» پیداست، روش «گرادیان نزولی» در حذف اثر نویز با دامنهٔ بالا[2] در بیشتر موارد نسبت به روش «الگوریتم خوشه‌بندی K-Means» بهتر عمل می‌کند. البته این بهتر بودن به قیمت حجم بالای محاسباتی که منجر به افزایش زمان اجرای برنامه می‌گردد به دست آمده است.

زمان اجرای برنامه، وابسته به الف) تعداد پارامترهای انتخاب شده برای آموزش ب) حجم محاسبات ج) برخی ویژگی‌های روش آموزش مانند پیچیدگی الگوریتم آن؛ می‌باشد. به عنوان مثال با افزایش تعداد نُرون‌های لایهٔ پنهان، تعداد پارامترهای انتخاب شده برای آموزش و در نتیجه زمان اجرای برنامه بیشتر می‌گردد. هم‌چنین محاسبات بازه‌ای در لایهٔ پنهان و لایهٔ خروجی به شدّت بر حجم محاسبات می‌افزاید. این قبیل محاسبات در روش «گرادیان نزولی» حضور پر رنگتری دارند. و یا به دلیل کمتر بودن پیچیدگی «الگوریتم خوشه‌بندی K-Means» نسبت به «گرادیان نزولی»، آموزش شبکه به کمک این الگوریتم؛ زمان اجرای کمتری دارد. در جدول۷، نسبت زمان اجرای برنامه با روش «الگوریتم خوشه‌بندی K-Means» به زمان اجرای برنامه با روش «گرادیان نزولی» برای تعداد متفاوتی از نُرون‌های لایهٔ پنهان آورده شده است.

جدول۷. نسبت زمان اجرای برنامه با روش «الگوریتم خوشه‌بندی K-Means» به زمان اجرای برنامه با روش «گرادیان نزولی» (زمان‌ها بر حسب ثانیه است)

تعداد نُرون‌های لایهٔ پنهان	سیستم غیر خطی پویای U شکل با پنج ورودی	سری زمانی مکی‌گلاس
۲۰۰	$\frac{41.6}{120.1} = 0.35$	$\frac{30}{73.4} = 0.41$
۲۰	$\frac{14.2}{37.5} = 0.38$	$\frac{11.5}{26.7} = 0.43$
۱۰	$\frac{13.1}{33.4} = 0.39$	$\frac{10}{22.5} = 0.44$

۶- نتیجه‌گیری

همانطور که از نتایج معلوم می‌گردد صرفنظر از روش آموزش؛ عملکرد شبکهٔ عصبی RBF گرانولی نسبت به شبکهٔ عصبی RBF در شرایطی که داده‌های ورودی آغشته به نویز هستند بهتر می‌باشد. هم‌چنین مشاهده می‌گردد که هر چه میزان نویز تزریق شده به داده‌ها بیشتر باشد درصد بهبود عملکرد شبکهٔ عصبی RBF گرانولی نسبت به شبکهٔ عصبی RBF افزایش می‌یابد. به عبارت دیگر بازه‌ای بودن پارامترها سیستم را در مقابل نویز، قوی[3] می‌گرداند. این مسأله بدان سبب است که تابع فعال‌ساز گرانولی در ساختار شبکهٔ عصبی RBF، با تنظیم «پارامترهای مربوط به پهنای عدم قطعیّت[4]، یعنی کران پائین و کران بالای انحراف استاندارد، باعث کاهش حساسیت به تغییرات ورودی می‌شود. بنابراین توصیه می‌گردد که اگر نویز زیادی در داده‌ها وجود دارد با توجه به بهتر بودن عملکرد شبکهٔ عصبی RBF گرانولی نسبت به شبکهٔ عصبی RBF، از شبکهٔ عصبی RBF گرانولی استفاده شود. اما اگر نویز موجود؛ مقدار قابل توجهی نداشته باشد، با توجه به سادگی ساختار شبکهٔ عصبی RBF از این ساختار استفاده گردد. سادگی ساختار شبکهٔ عصبی RBF که به دلیل رهایی از محاسبات بازه‌ای در لایه‌های میانی و خروجی می‌باشد، سبب کم شدن خطای ناشی از این محاسبات و هم‌چنین کاهش حجم محاسباتی برنامه می‌گردد. کم شدن مقدار تقریب در محاسبهٔ خروجی، می‌تواند در شرایط بدون نویز و به ویژه در شناسایی سیستم‌های غیر خطی پویا؛ منجر به اندکی بهبود در نتایج گردد.

از طرفی نتایج به دست آمده از شبکهٔ عصبی RBF گرانولی، در شرایطی که داده‌های ورودی آغشته به نویز قوی نیستند؛ نشان دهندهٔ این است که عملکرد «الگوریتم خوشه‌بندی K-Means» در شناسایی سیستم‌های غیر خطی پویا که آشوبی نیستند بهتر است. البته برای شناسایی و پیش‌بینی توابع آشوب برای افق زمانی بزرگ، نتایج به دست آمده بیانگر آن است که عملکرد شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیلهٔ «الگوریتم گرادیان نزولی» عملکرد بهتری دارد. هر چند که «الگوریتم خوشه‌بندی K-Means» نیز در این شرایط با کیفیّت پایین‌تری کاربرد دارد.

با افزایش قدرت نویز در داده‌های ورودی برتری «الگوریتم گرادیان نزولی» به تدریج افزایش می‌یابد تا جایی که در نویز با «نسبت سیگنال به نویز» برابر صفر، عملکرد شبکهٔ عصبی RBF گرانولی با آموزش پارامترهای لایهٔ میانی بوسیله‌ی روش «گرادیان نزولی»؛ برای شناسایی سیستم‌های غیر خطی پویا و شناسایی و پیش‌بینی توابع آشوب، بهتر از حالتی است که پارامترهای لایهٔ میانی بوسیلهٔ روش «الگوریتم خوشه‌بندی K-Means» آموزش می‌بینند. چرا که در هنگام وجود نویز قوی در دادهٔ ورودی، تعیین تعلق آن به یک خوشه دشوار می‌باشد و خطای بوجود آمده در تعیین خوشه بر تقریب ناشی از محاسبات بازه‌ای غلبه می‌یابد.

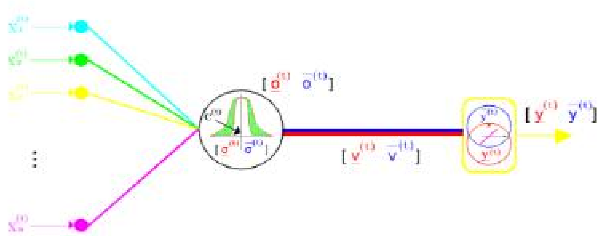
با این وجود، زمان اجرای برنامه با آموزش پارامترهای لایهٔ میانی بوسیلهٔ «الگوریتم خوشه‌بندی K-Means»، کمتر از روش گرادیان نزولی است. علت آن این است که مرتبهٔ پیچیدگی الگوریتم و حجم محاسباتی برنامه کاهش یافته و همگرایی پارامترها سریع‌تر صورت می‌پذیرد. از اینرو در شرایطی که داده‌های ورودی آغشته به نویز قوی نیستند؛ در تعداد مرحله مساوی، نسبت به روش «گرادیان نزولی»؛ به جواب بهتری خواهیم رسید.

[1] Run time
[2] High Amplitude
[3] Robust
[4] The parameters responsible for the width of uncertainly

۷- پیوست:اثبات بهبود یافتن عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF در شرایط نویزی

۱-۷ اثبات برای حالتی که یک نُرون در لایهٔ میانی وجود داشته باشد

برای اثبات بهبود یافتن عملکرد شبکهٔ عصبیِ RBF گرانولی نسبت به شبکهٔ عصبیِ RBF در شرایط نویزی؛ از یک شبکهٔ عصبیِ RBF گرانولی با یک نُرون در لایهٔ میانی استفاده می‌کنیم. اثبات می‌گردد که هر چه میزان نویز تزریق شده به داده‌ها بیشتر باشد، شبکهٔ عصبیِ RBF گرانولی با افزایش «فاصله کران پائین انحرافِ استاندارد از کران بالای آن» و در نتیجه افزایش «بهنایِ عدم قطعیت و یا همان تابع فعال‌ساز گاوسیِ گرانولی»، اثر نویز در خروجی شبکه را مهار می‌نماید. هم‌چنین هر چه میزان نویز تزریق شده به داده‌ها کمتر باشد ساختار شبکهٔ عصبیِ RBF گرانولی به شبکهٔ عصبیِ RBF نزدیکتر می‌شود تا در شرایط بدون نویز دو ساختار مشابه یکدیگر می‌گردند. البته این تغییر ساختار ناشی از بازه‌ای بودن دو پارامترِ سیستم، یکی «انحرافِ استاندارد تابع فعال‌ساز گرانولی» در لایهٔ میانی؛ و دیگری «وزن‌های بازه‌ای» در لایهٔ خروجی؛ می‌باشد.



شکل ۹. یک شبکهٔ عصبیِ RBF گرانولی با یک نُرون در لایهٔ میانی

محاسبه کرانِ پائینِ خروجیِ یک تابعِ گاوسیِ گرانولی در لایهٔ میانی به شکل زیر است:

$$\underline{o}^{(t)} = \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{(t)}}{\underline{\sigma}_r^{(t)}}\right)^2\right] = \psi\left[\sum_{r=1}^{n}\left(\overline{u}^{(t)}\right)^2\right] = \psi\left(\overline{d}^{(t)}\right) = exp\left(-\frac{1}{2}\overline{d}^{(t)}\right) \quad (۷۹)$$

اکنون برای محاسبه کرانِ بالایِ خروجیِ یک تابعِ گاوسیِ گرانولی در لایهٔ میانی خواهیم داشت:

$$\bar{o}^{(t)} = \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{(t)}}{\overline{\sigma}_r^{(t)}}\right)^2\right] = \psi\left[\sum_{r=1}^{n}\left(\underline{u}_r^{(t)}\right)^2\right] = \psi\left(\underline{d}^{(t)}\right) = exp\left(-\frac{1}{2}\underline{d}^{(t)}\right) \quad (۸۰)$$

همان‌طور که در محاسبه خروجیِ نهاییِ شبکهٔ عصبیِ RBF گرانولی دیدیم؛ برای محاسبه خروجیِ نهاییِ این شبکه با یک نُرون هم چهار حالت زیر را خواهیم داشت.

- حالت اول: اگر $v^{(t)} > 0$ و $v^{(t)} > |s^{(t)}|$ باشد آن‌گاه $\underline{v}^{(t)}, \overline{v}^{(t)} > 0$ خواهد بود.
- حالت دوم: اگر $v^{(t)} > 0$ و $v^{(t)} < |s^{(t)}|$ باشد آن‌گاه $\underline{v}^{(t)} < 0$ و $\overline{v}^{(t)} > 0$ خواهد بود.
- حالت سوم: اگر $v^{(t)} < 0$ و $|v^{(t)}| > |s^{(t)}|$ باشد آن‌گاه $\underline{v}^{(t)}, \overline{v}^{(t)} < 0$ خواهد بود.
- حالت چهارم: اگر $v^{(t)} < 0$ و $|v^{(t)}| < |s^{(t)}|$ باشد آن‌گاه $\underline{v}^{(t)} < 0$ و $\overline{v}^{(t)} > 0$ خواهد بود.

در حالت‌های اول، دوم و چهارم، کرانِ پایین و کرانِ بالایِ خروجیِ نهاییِ شبکهٔ عصبیِ RBF گرانولی با یک نُرون، به همراه پارامترهای خطی کرانِ پایین و کرانِ بالایِ لایهٔ خروجیِ شبکه به این صورت خواهد بود:

$$\underline{y_{\mathcal{N}}}^{(t)} = \underline{v}^{(t)} \times \underline{o}^{(t)} = \left(v^{(t)} - |s^{(t)}|\right) \times \underline{o}^{(t)} \quad (۸۱)$$

$$\overline{y_{\mathcal{N}}}^{(t)} = \overline{v}^{(t)} \times \overline{o}^{(t)} = \left(v^{(t)} + |s^{(t)}|\right) \times \overline{o}^{(t)} \quad (۸۲)$$

در حالت سوم به علت منفی بودن $\underline{v}^{(t)}$ و $\overline{v}^{(t)}$ و مثبت بودن $\underline{o}^{(t)}$ و $\overline{o}^{(t)}$، محاسبه کرانِ پایین و کرانِ بالایِ خروجیِ نهاییِ شبکهٔ عصبیِ RBF گرانولی با یک نُرون، به همراه پارامترهای خطی کرانِ پایین و کرانِ بالایِ لایهٔ خروجیِ شبکه به این صورت خواهد بود:

$$\underline{y_{\mathcal{N}}}^{(t)} = \underline{v}^{(t)} \times \underline{o}^{(t)} = \left(v^{(t)} - |s^{(t)}|\right) \times \overline{o}^{(t)} \quad (۸۳)$$

$$\overline{y_{\mathcal{N}}}^{(t)} = \overline{v}^{(t)} \times \overline{o}^{(t)} = \left(v^{(t)} + |s^{(t)}|\right) \times \underline{o}^{(t)} \quad (۸۴)$$

و خروجی نهایی از میانگین کرانِ پائین و کرانِ بالا به دست می‌آید:

$$y_{\mathcal{N}}^{(t)} = \frac{\underline{y_{\mathcal{N}}}^{(t)} + \overline{y_{\mathcal{N}}}^{(t)}}{2} \quad (۸۵)$$

اکنون با سه فرض زیر اثبات را ارائه می کنیم:

- وزنِ لایهٔ خروجی، یعنی $v^{(t)}$ بازه‌ای نبوده و مساوی یک باشد، در آن صورت خواهیم داشت:

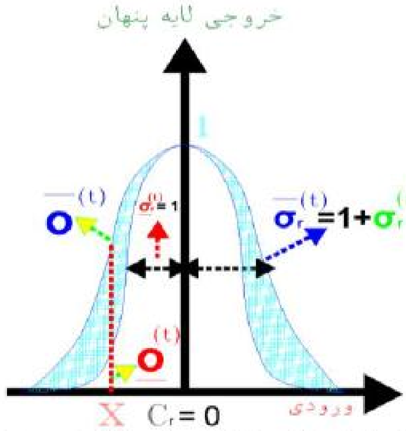
$$y_{\mathcal{N}}^{(t)} = \frac{\underline{y_{\mathcal{N}}}^{(t)} + \overline{y_{\mathcal{N}}}^{(t)}}{2} = \frac{\left(\overline{v}^{(t)} \times \overline{o}^{(t)}\right) + \left(\underline{v}^{(t)} \times \underline{o}^{(t)}\right)}{2} = \frac{\overline{o}^{(t)} + \underline{o}^{(t)}}{2} = \frac{\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{(t)}}{\underline{\sigma}_r^{(t)}}\right)^2\right] + \psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}-c_r^{(t)}}{\overline{\sigma}_r^{(t)}}\right)^2\right]}{2} \quad (۸۶)$$

- هم‌چنین مقدار اولیه‌ی مرکز تابعِ فعال‌سازِ گاوسیِ نُرونِ لایهٔ پنهان یعنی c مساوی صفر در نظر گرفته می‌شود.

$$y_{\mathcal{N}}^{(t)} = \frac{\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}}{\underline{\sigma}_r^{(t)}}\right)^2\right]+\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}}{\overline{\sigma}_r^{(t)}}\right)^2\right]}{2} \quad (۸۷)$$

- کران پائینِ انحراف استاندارد مساوی یک ($\underline{\sigma}_r^{(t)} = 1$) و کرانِ بالایِ انحراف استاندارد به صورت $\overline{\sigma}_r^{(t)} = 1 + \sigma_r^{(t)}$ در نظر گرفته می‌شود که $\sigma_r^{(t)} = [0 \quad 3]$ بازهٔ تغییرات انحراف استاندارد با گام ۰/۰۱ خواهد بود.

$$y_{\mathcal{N}}^{(t)} = \frac{\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}}{1}\right)^2\right]+\psi\left[\sum_{r=1}^{n}\left(\frac{x_r^{(t)}+n_r^{(t)}}{1+\sigma_r^{(t)}}\right)^2\right]}{2} = \frac{1}{2}\times \psi\left[\sum_{r=1}^{n}\left(x_r^{(t)}+n_r^{(t)}\right)^2\right] + \frac{1}{2}\times\psi\left[\sum_{r=1}^{n}\frac{\left(x_r^{(t)}+n_r^{(t)}\right)^2}{\left(1+\sigma_r^{(t)}\right)^2}\right] \quad (۸۸)$$



شکل ۱۰. تابع فعال‌ساز گاوسیِ گرانولی در نُرون لایهٔ میانی و پیش از آغاز آموزش مرکز تابع فعال‌ساز

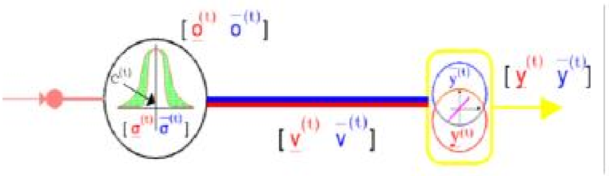
سپس خروجی شبکه بدون افزودن نویز ورودی و با فرضیات بالا و در نظر گرفتن بردار X به عنوان متغیر مستقل به شکل زیر به دست می‌آید:

$$y^{(t)} = f_{(t)}(\boldsymbol{X}) = f\left(x_1^{(t)}, x_2^{(t)}, x_3^{(t)}, \cdots x_n^{(t)}\right) = \frac{1}{2}\times\psi\left[\sum_{r=1}^{n}\left(x_r^{(t)}\right)^2\right] + \frac{1}{2} \times \psi\left[\sum_{r=1}^{n}\frac{\left(x_r^{(t)}\right)^2}{\left(1+\sigma_r^{(t)}\right)^2}\right] \quad (۸۹)$$

با افزودن نویز به داده‌ها، معادله خروجی به صورت زیر خواهد بود:

$$y_{\mathcal{N}}^{(t)} = f_{(t)}(\boldsymbol{X},\boldsymbol{\mathcal{N}}) = f\left(x_1^{(t)}, x_2^{(t)}, x_3^{(t)}, \cdots x_n^{(t)}, n_1^{(t)}, n_2^{(t)}, n_3^{(t)}, \cdots, n_n^{(t)}\right) = \frac{1}{2}\times\psi\left[\sum_{r=1}^{n}\left(x_r^{(t)}+n_r^{(t)}\right)^2\right] + \frac{1}{2} \times\psi\left[\sum_{r=1}^{n}\frac{\left(x_r^{(t)}+n_r^{(t)}\right)^2}{\left(1+\sigma_r^{(t)}\right)^2}\right] \quad (۹۰)$$

اکنون برای کاهش تعداد متغیرها در تابع خروجی در وضعیت بدون نویز و نویزی، با فرض این‌که ابعاد بردار ورودی برابر یک باشد یعنی $n = 1$، اثبات را ادامه می‌دهیم:



شکل ۱۱. یک شبکهٔ عصبی RBF گرانولی با یک نُرون در لایهٔ میانی و بردار یک بعدی ورودی

در این صورت خواهیم داشت:

$$y^{(t)} = f\left(x^{(t)}\right) = \frac{1}{2}\times\psi\left[\left(x^{(t)}\right)^2\right] + \frac{1}{2}\times\psi\left[\frac{\left(x^{(t)}\right)^2}{\left(1+\sigma^{(t)}\right)^2}\right] \quad (۹۱)$$

$$y_{\mathcal{N}}^{(t)} = f\left(x^{(t)}, n^{(t)}\right) = \frac{1}{2}\times\psi\left[\left(x^{(t)}+n^{(t)}\right)^2\right] + \frac{1}{2}\times\psi\left[\frac{\left(x^{(t)}+n^{(t)}\right)^2}{\left(1+\sigma^{(t)}\right)^2}\right] \quad (۹۲)$$

تفاضل این دو تابع، تابعی است که اختلال ایجاد شده بوسیلهٔ نویز[1] را نشان می‌دهد و عبارت است از:

$$DCN = \left|y_{\mathcal{N}}^{(t)} - y^{(t)}\right| = \frac{1}{2} \times \left|\psi\left[\left(x^{(t)}+n^{(t)}\right)^2\right] + \psi\left[\frac{\left(x^{(t)}+n^{(t)}\right)^2}{\left(1+\sigma^{(t)}\right)^2}\right] - \psi\left[\left(x^{(t)}\right)^2\right] - \psi\left[\frac{\left(x^{(t)}\right)^2}{\left(1+\sigma^{(t)}\right)^2}\right]\right| \quad (۹۳)$$

$$DCN = \frac{1}{2}\times\left|exp\left[\frac{-\left(x^{(t)}+n^{(t)}\right)^2}{2}\right] + exp\left[\frac{-\left(x^{(t)}+n^{(t)}\right)^2}{2\left(1+\sigma^{(t)}\right)^2}\right] - exp\left[\frac{-\left(x^{(t)}\right)^2}{2}\right] - exp\left[\frac{-\left(x^{(t)}\right)^2}{2\left(1+\sigma^{(t)}\right)^2}\right]\right| \quad (۹۴)$$

همان‌طور که در بخش ۲ گفته شد، تابع توزیع احتمال گاوسی یا نرمال[2] بر اساس قضیه حد مرکزی از اهمیّت زیادی برخوردار است. اکنون «تابع توزیع احتمال» برای نویز ورودی را که به آن «تابع توزیع

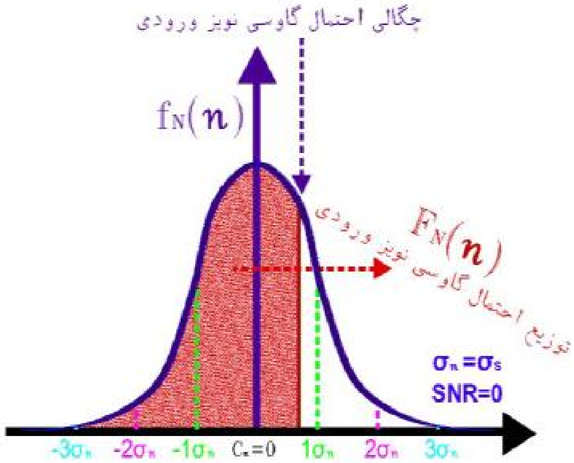
[1] Distortion Caused by Noise (DCN)
[2] Gaussian Probability Distribution Function (Gaussian PDF)

تجمعی[1]، نیز گفته می‌شود، از نوع «نرمال یا گاوسی با میانگین صفر» و به صورت $\mathcal{N}(0,\sigma_n^2)$ در نظر می‌گیریم و به شکل زیر تعریف می‌نمائیم[۳۹ ۴۱]:

$$F_{\mathcal{N}}(n)=\int_{-\infty}^{n} f_{\mathcal{N}}(z)\,dz=\frac{1}{\sqrt{2\pi}\times\sigma_n}\int_{-\infty}^{n}\psi\left(\frac{z^2}{\sigma_z^2}\right)dz=\frac{1}{\sqrt{2\pi}\times\sigma_n}\int_{-\infty}^{n}exp\left(-\frac{1}{2}\times\frac{z^2}{\sigma_z^2}\right)dz \qquad (۹۵)$$

تابع چگالی احتمال[2] برای نویز ورودی به شکل زیر به دست می‌آید.

$$f_{\mathcal{N}}(n)=\frac{d}{dn}F_{\mathcal{N}}(n)=\frac{1}{\sqrt{2\pi}\times\sigma_n}\psi\left(\frac{n^2}{\sigma_n^2}\right)=\frac{1}{\sqrt{2\pi}\times\sigma_n}exp\left(-\frac{1}{2}\times\frac{n^2}{\sigma_n^2}\right) \qquad (۹۶)$$



شکل۱۲. تابع توزیع احتمال و تابع چگالی احتمال گاوسی برای نویز ورودی

همان طور که در شکل۹ می‌بینیم، فرض بر این است که قدرت نویز زیاد بوده و با قدرت سیگنال برابر باشد یعنی خواهیم داشت $SNR=0\ (dB)$ و در نتیجه [۱۱]:

$$SNR=10\log_{10}\frac{\text{the variance of the signal}}{\text{the variance of the noise}}=10\log_{10}\frac{\sigma_s^2}{\sigma_n^2}=0\rightarrow\sigma_s=\sigma_n \qquad (۹۷)$$

آنجا که σ_s^2 واریانس سیگنال و σ_n^2 واریانس نویز می‌باشد. اکنون فرض می‌کنیم $\sigma_s=\sigma_n=\sigma'=0.33$ باشد، در آن صورت تابع چگالی احتمال برای نویز که در حالت گسسته به آن تابع جرم احتمال[3] گویند به شکل زیر ساده می‌گردد:

$$f_{\mathcal{N}}(n)=\frac{1}{\sqrt{2\pi}\times\sigma'}exp\left(-\frac{1}{2}\times\frac{n^2}{\sigma'^2}\right)=\frac{1}{2.5\times0.33}e^{-\frac{n^2}{2\times0.33^2}}=\frac{1}{0.825}e^{-\frac{n^2}{0.2178}}=1.2e^{-4.6n^2} \qquad (۹۸)$$

اکنون با فرض داشتن یک نمونه ورودی یعنی $T=1$، می‌توانیم تابع دو متغیره DCN را به یک تابع یک متغیره با متغیر مستقل n تبدیل نمائیم.

$$DCN=|y_{\mathcal{N}}-y|=\frac{1}{2}\times\left|exp\left[\frac{-(x+n)^2}{2}\right]+exp\left[\frac{-(x+n)^2}{2(1+\sigma)^2}\right]-exp\left[\frac{-x^2}{2}\right]-exp\left[\frac{-x^2}{2(1+\sigma)^2}\right]\right| \qquad (۹۹)$$

متغیر تصادفی n روی فضای احتمال[4] $(\Omega,\mathcal{F},P)$ تعریف می‌گردد که در آن داریم[۴۱]:

- Ω فضای نمونه[5] شامل مقادیر تصادفی نویز و در واقع همان نتایج به دست آمده از آزمایشات است.
- $\mathcal{F}$ فضای رویداد[6] و یا فضای فرآیند تصادفی[7] است که بخشی از فضای نمونه را در بر می‌گیرد و شامل رویدادهایی است که وقوع پذیرند و در واقع احتمال وقوع بزرگتر از صفر دارند. به عنوان مثال رویداد قرار گرفتن «بزرگی نویز در بازهٔ $[-0.6745\sigma_n \quad 0.6745\sigma_n]$»، احتمال وقوع بالایی دارد.
- P تابع اندازهٔ احتمال رویداد[8] است که وظیفه آن اختصاص یک عدد حقیقی به رویداد می‌باشد و در واقع فضای رویداد را به فضای اعداد حقیقی نگاشت می‌کند. می‌توان از تابع چگالی احتمال به عنوان تابع اندازهٔ احتمال رویداد استفاده نمود.

مقدار مورد انتظار یا اُمید ریاضی برای اختلال ایجاد شده بوسیله نویز، عبارت است از:

$$E(\mathcal{N})=\int_{\Omega}DCN\times dP=\int_{-\infty}^{\infty}DCN\times f_{\mathcal{N}}(n)dn=\int_{-\infty}^{\infty}|y_{\mathcal{N}}-y|\times f_{\mathcal{N}}(n)dn \qquad (۱۰۰)$$

با تقریب خوبی می‌توان حدود انتگرال‌گیری را به شکل زیر عوض نمود:

$$E(\mathcal{N})=\int_{-4\sigma}^{4\sigma}DCN\times f_{\mathcal{N}}(n)dn=\int_{-4\sigma}^{4\sigma}|y_{\mathcal{N}}-y|\times f_{\mathcal{N}}(n)dn \qquad (۱۰۱)$$

می‌توان فرم پیوسته تابع اُمید ریاضی یعنی $E(\mathcal{N})$ را به صورت زیر گسسته نمود. در حالت گسسته این تابع از طریق جمع مقادیر نمونه‌ها و

[4] Probability space
[5] Sample space
[6] Event space
[7] Stochastic Process space
[8] Event Probability Function

[1] Cumulative Distribution Function (CDF)
[2] Probability Density Function (pdf)
[3] Probability Mass Function (pmf)

هنجارسازی[1] حاصل جمع نسبت به تعداد نمونه‌ها محاسبه می‌شود و یا می‌توان از تابع چگالی احتمال نویز ورودی به صورت زیر استفاده نمود.

$$E(\mathcal{N}) = \sum_{n=-4\sigma}^{n=4\sigma}(|y_{\mathcal{N}} - y| \times f_{\mathcal{N}}(n)\Delta n) \qquad (۱۰۲)$$

هر چه مقدار Δn را کوچکتر در نظر بگیریم دقت محاسبه $E(\mathcal{N})$ بیشتر خواهد بود. Δn را می‌توان برابر ۰/۰۰۱ یا ۰/۰۰۰۱ در نظر گرفت. اکنون تابع $E(\mathcal{N}^2)$ را به شکل زیر به دست می‌آوریم:

$$E(\mathcal{N}^2) = \int_{-4\sigma}^{4\sigma}(DCN)^2 \times f_{\mathcal{N}}(n)dn = \int_{-4\sigma}^{4\sigma}(y_{\mathcal{N}} - y)^2 \times f_{\mathcal{N}}(n)dn \qquad (۱۰۳)$$

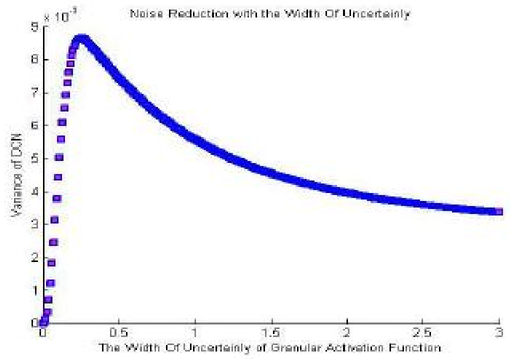
می‌توان فرم پیوسته تابع $E(\mathcal{N}^2)$ را به صورت زیر گسسته نمود:

$$E(\mathcal{N}^2) = \sum_{n=-4\sigma}^{n=4\sigma}[(y_{\mathcal{N}} - y)^2 \times f_{\mathcal{N}}(n)\Delta n] \qquad (۱۰۴)$$

واریانسِ تابع DCN به شکل زیر محاسبه می‌گردد:

$$Var(\mathcal{N}) = E(\mathcal{N}^2) - [E(\mathcal{N})]^2 \qquad (۱۰۵)$$

نمودار تغییرات واریانس تابع DCN بر حسب σ مانند شکل ۱۰ به دست می‌آید. این شکل نشان می‌دهد که ابتدا با افزایش مقدار σ (تغییرِ انحراف استاندارد تابع فعال‌سازِ گرانولی) تا سطح $\sigma' = 0.33$ (انحراف استاندارد نویز)، واریانسِ تابع DCN افزایش می‌یابد. چرا که پهنای عدم قطعیت در تابع فعال‌سازِ گرانولی هنوز قدرت خنثی نمودن اثر نویز را ندارد و تغییرات مقدارنویز سبب تغییر اختلال ایجاد شده بوسیله نویز می‌گردد. که آن نیز به نوبه خود تغییر در واریانس این اختلال را در پی دارد که در این مرحله افزایشی می‌باشد. با افزایش بیشتر مقدار σ، پهنای عدم قطعیت در تابع فعال‌سازِ گرانولی بر نویز ورودی غلبه نموده و واریانس تابع DCN رو به کاهش می‌گذارد. کاهش واریانسِ تابع DCN به معنی آن است که با افزایش پهنای عدم قطعیت در تابع فعال‌سازِ گرانولی، مقدار تابع DCN حول میانگین آن پراکندگی کمتری دارد و چون تابع توزیع احتمال نویز ورودی، گاوسی یا نُرمال با میانگین صفر می‌باشد، پس تابع DCN بیشتر به سمت صفر میل کرده و در واقع اختلال ایجاد شده بوسیله نویز کاهش یافته است.



شکل۱۳. تغییرات واریانسِ تابع DCN بر حسب پهنای عدم قطعیت یعنی σ

۷-۲ تعمیم اثبات فوق برای حالتی که m نُرون در لایهٔ میانی وجود داشته باشد

تعریف ۱. فرآیندِ تصادفیِ شمارا[2] (با زمان گسسته[3]) روی فضای احتمال $(\Omega, \mathcal{F}, P)$ عبارت است از دنباله‌ای از متغیرهای تصادفی مانند $\left(n_r^{(t)}\right)_{r\in N}$ روی این فضا و آن را به صورت $\left\{n_r^{(t)}\right\}_{r\in N}$ نشان می‌دهیم[۳۹].

قرارداد. از این پس منظور از $\mathbb{R}^\infty$ یعنی مجموعهٔ همهٔ دنباله‌های اعداد حقیقی، و یا به عبارت دیگر مجموعهٔ همهٔ توابع از $\mathbb{N}$ به $\mathbb{R}$ که هر تابع می‌تواند یک دنباله را تولید کند. به عنوان مثال تابع نویز $n_r^{(t)} = \neq(r,t)$ با «تابع توزیع احتمالِ گاوسی» می‌تواند یکی از این توابع باشد[۳۸، ۴۰].

تعریف ۲. مجموعهٔ A در $\mathbb{R}^\infty$ را یک بازه با بُعد متناهی گوییم اگر $n \in N$ و بازه‌های متناهی و نامتناهی I_1، I_2، I_3، ...، I_n موجود باشند که $A = \left\{\mathcal{N}^{(t)} \in \mathbb{R}^\infty\colon\ n_r^{(t)} \in I_r,\ \ r = 1, 2, \ldots, n\right\}$. به عنوان مثال مقدار نویز افزوده شده به ورودی r اُم، یعنی $x_r^{(t)}$ که به صورت $n_r^{(t)}$ نشان داده می‌شود، در بازهٔ $[-4\sigma_n \quad 4\sigma_n] \in n_r^{(t)}$ که یک بازه با بُعد متناهی است قرار دارد. بُعد متناهی این بازه برابر بُعدِ بردار نویز ورودی یعنی $\mathcal{N}^{(t)} = \left[n_1^{(t)}\ n_2^{(t)}\ \ldots\ n_n^{(t)}\right]$ می‌باشد. ردهٔ تمام بازه‌های با بُعد متناهی را با $\mathcal{G}_\infty$ نشان می‌دهند[۲۷].

تعریف ۳. فرض می‌کنیم که مجموعهٔ B به صورت زیر تعریف گردد[۳۸، ۴۰]:

$$B = \{X^{(t)}, \mathcal{N}^{(t)} \in \mathbb{R}^\infty\colon\ (x_1^{(t)}, x_2^{(t)}, x_3^{(t)}, \cdots x_n^{(t)}, n_1^{(t)}, n_2^{(t)}, n_3^{(t)}, \cdots, n_n^{(t)}) \in B_n\} \qquad (۱۰۶)$$

- مجموعهٔ B در $\mathbb{R}^\infty$ را یک استوانه[4] با بُعد متناهی گوییم.
- ردهٔ تمام استوانه‌های با بُعد متناهی با e_∞ نشان داده می‌شود که e_∞ یک میدان است.
- در این صورت e_∞ می‌تواند یک میدان سیگمایی[5] باشد.
- از دیدگاه احتمال، میدان سیگمایی را می‌توان به عنوان اطلاعات داده شدهٔ یک آزمایش در نظر گرفت.
- مجموعه‌های بُرِل روی Ω و P از فضای احتمال $(\Omega, \mathcal{F}, P)$ به شکل $\sigma(e_\infty)$ و $\sigma(\mathcal{G}_\infty)$ تعریف می‌گردند.

[1] Normalizing

[2] Countable Stochstic Process
[3] Discrete Time
[4] Cylinder
[5] σ-field

- مجموعهٔ B_∞ را به صورت $B_\infty = \sigma(\mathcal{G}_\infty) = \sigma(e_\infty)$ تعریف می‌کنیم.

گزاره: اثبات می‌گردد که هر فرآیند تصادفی مانند فرآیندِ تصادفیِ شمارای $\{n_r^{(t)}\}_{r\epsilon N}$، یک بردار تصادفی به صورت $\mathcal{N}^{(t)}:(\Omega,\mathcal{F},P)\to(\mathbb{R}^\infty,B_\infty)$ به دست می‌دهد[۳۹ ۴۱].

قضیهٔ توسیع کُلموگُرف[1]: فرض کنید که برای هر $n\in N$ یک اندازهٔ احتمال P_n روی $(\mathbb{R}^\infty,B)$ در دست است. می‌خواهیم یک اندازهٔ احتمال P روی $(\mathbb{R}^\infty,B_\infty)$ طوری بسازیم که تحدید P به B_n همان P_n شود. یعنی داشته باشیم[۳۹ ۴۱]:

$$\forall B\epsilon B_n;\quad P\left(\left\{\boldsymbol{X}^{(t)},\boldsymbol{\mathcal{N}}^{(t)} \in \mathbb{R}^\infty:\ \left(x_1^{(t)},x_2^{(t)},x_3^{(t)},\cdots x_n^{(t)},n_1^{(t)},n_2^{(t)},n_3^{(t)},\cdots,n_n^{(t)}\right) \in B\right\}\right)=P_n(B) \tag{۱۰۷}$$

می‌دانیم که $\mathcal{F}\subset\Omega$ است. اگر مجموعهٔ F را به عنوان مجموعه‌ای از درایه‌های بردار $\mathcal{N}^{(t)}$ در نظر بگیریم و F یک مجموعهٔ شمارا و زیر مجموعهٔ R باشد. و با توجه به این‌که می‌دانیم، دنبالهٔ $\left(n_r^{(t)}\right)_{r\epsilon N}$ دنباله‌ای از متغیرهای تصادفی است آن‌گاه در صورت برقراری شرایط زیر:

- $P(.)\geq 0,$ (۱۰۸)

- $\sum_{n_{n+1}^{(t)}\epsilon F} P\left(n_1^{(t)},n_2^{(t)},n_3^{(t)},\cdots,n_{n+1}^{(t)}\right)=P\left(n_1^{(t)},n_2^{(t)},n_3^{(t)},\cdots,n_n^{(t)}\right),$ (۱۰۹)

- $\sum_{n_1^{(t)}\epsilon F} P\left(n_1^{(t)}\right)=1,$ (۱۱۰)

فرآیندِ شمارای $\{x_r^{(t)}\}_{r\epsilon N}$ که می‌تواند دارای ویژگی غیرخطی و یا آشوبی باشد، چنان موجود است که:

$$P\left(x_1^{(t)}\leftarrow x_1^{(t)}+n_1^{(t)},\ldots,x_n^{(t)}\leftarrow x_n^{(t)}+n_n^{(t)}\right)=P\left(n_1^{(t)},n_2^{(t)},n_3^{(t)},\cdots,n_n^{(t)}\right) \tag{۱۱۱}$$

برای تعمیم تعداد ورودی‌ها از یک ورودی به چند ورودی، از قضیهٔ توسیع کُلموگُرف و اثبات به کمک استقراءِ ریاضی، به شکل زیر استفاده می‌کنیم:

- اگر برای $1\in N$ یک اندازهٔ احتمال P_1 روی $(\mathbb{R}^\infty,B)$ در دست داشته باشیم، می‌توانیم یک اندازهٔ احتمال P روی $(\mathbb{R}^\infty,B_\infty)$ طوری بسازیم که تحدید P به B_1 همان P_1 شود. یعنی داشته باشیم:

$$\forall B\epsilon B_1;\quad P\left(\left\{\boldsymbol{X}^{(t)},\boldsymbol{\mathcal{N}}^{(t)} \in \mathbb{R}^\infty:\ \left(x_1^{(t)},n_1^{(t)}\right)\in B\right\}\right)=P_1(B) \tag{۱۱۲}$$

- اگر برای $r\in N$ یک اندازهٔ احتمال P_r روی $(\mathbb{R}^\infty,B)$ در دست داشته باشیم، می‌توانیم یک اندازهٔ احتمال P روی $(\mathbb{R}^\infty,B_\infty)$ طوری بسازیم که تحدید P به B_r همان P_r شود. یعنی داشته باشیم:

$$\forall B\epsilon B_r;\quad P\left(\left\{\boldsymbol{X}^{(t)},\boldsymbol{\mathcal{N}}^{(t)} \in \mathbb{R}^\infty:\ \left(x_1^{(t)},x_2^{(t)},n_1^{(t)},n_2^{(t)}\right) \in B\right\}\right)=P_r(B) \tag{۱۱۳}$$

- اگر برای $r+1\in N$ یک اندازهٔ احتمال P_{r+1} روی $(\mathbb{R}^\infty,B)$ در دست داشته باشیم، می‌توانیم یک اندازهٔ احتمال P روی $(\mathbb{R}^\infty,B_\infty)$ طوری بسازیم که تحدید P به B_{r+1} همان P_{r+1} شود. یعنی داشته باشیم:

$$\forall B\epsilon B_{r+1};\quad P\left(\left\{\boldsymbol{X}^{(t)},\boldsymbol{\mathcal{N}}^{(t)} \in \mathbb{R}^\infty:\ \left(x_1^{(t)},x_2^{(t)},n_1^{(t)},n_2^{(t)}\right) \in B\right\}\right)=P_{r+1}(B) \tag{۱۱۴}$$

پس با هر تعداد ورودی قادر به ساختن یک اندازهٔ احتمال P روی $(\mathbb{R}^\infty,B_\infty)$ خواهیم بود.

قضیهٔ وجود نگاشت[2] شبکهٔ عصبیِ پیشرو[3] کُلموگُرف: هر تابع پیوستهٔ $f(x^{(t)},n^{(t)}):[0\quad 1]^n\to R^m$ را می‌توان به طور دقیق توسط یک شبکهٔ عصبی سه لایه‌ای پیشرو[4]، که دارای n نُرون در لایهٔ ورودی و $(2n+1)$ نُرون در لایهٔ پنهان و m نُرون در لایهٔ خروجی می‌باشد تحقق بخشید. طبق این قضیه چنین شبکه‌ای وجود دارد[۴۲].

قضیهٔ وجود نگاشت شبکهٔ عصبیِ پس‌انتشار[5] کُلموگُرف: برای هر تابع $f(x^{(t)},n^{(t)}):[0\quad 1]^n\to R^m$، که L_2 باشد یعنی انتگرال $\int_{[0\quad 1]^n}\left|f(x^{(t)},n^{(t)})\right|^2dx$ موجود باشد، یک شبکهٔ عصبی پس‌انتشار سه لایه وجود دارد که می‌تواند تابع f را با دقت میانگین مربع خطای یک عدد کسری ثابت مثبت به نام $0<e<d$ تقریب بزند.

فضای L_2 شامل هر تابعی که در یک مسألهٔ عملی وجود دارد می‌شود. به عنوان نمونه شامل تمام توابع ناپیوسته که به صورت تکه‌ای نیز پیوسته هستند می‌شود. این قضیه ثابت می‌کند که همیشه سه لایه کافیست. هر چند در عمل استفاده از ۴ یا ۵ و یا حتی لایه‌های بیشتر مرسوم است. زیرا بسیاری از مسائل برای حل، توسط شبکهٔ سه لایه‌ای، نیازمند تعداد زیادی نُرون در لایهٔ پنهان خود می‌باشند. که این امر به کند شدن سرعت پردازش منجر می‌شود. بالاخره هر چند این قضیه تضمین می‌کند که شبکه‌های عصبی چند لایه با وزن‌های صحیح، قادر به نگاشت

[1] Kolmogorov

[2] Mapping
[3] Feed Forward Neural Network (FFNN)
[4] Feed forward three layers neural network
[5] Back Propagation Neural Network (BPNN)

و تقریب هر تابع L_2 دلخواه است؛ ولی در مورد رسیدن و یا نرسیدن به این وزن‌ها، هیچ قانون یادگیری تضمینی موجود نیست[۴۲].

پس وقتی یک شبکهٔ عصبیِ پس‌انتشار سه لایه مانند شبکهٔ عصبیِ RBF گرانولی با یک نُرون در لایهٔ میانی و آموزش پارامترها به روش «گرادیان نزولی» بتواند با تقریب بهتری نسبت به یک شبکهٔ عصبیِ RBF با همان تعداد نُرون در لایهٔ میانی و همان روش آموزش، یک تابع غیرخطی یا آشوب را تحقق ببخشد، به طور حتم با چند نُرون در لایهٔ میانی نیز این برتری به وقوع خواهد پیوست. زیرا تابع غیرخطی و تابع آشوب به کار رفته، هر دو L_2 می‌باشند.

از طرفی دو قضیهٔ «وجود نگاشت شبکهٔ عصبیِ پیشرو کُلموگُرف» و «وجود نگاشت شبکهٔ عصبیِ پس‌انتشار کُلموگُرف» محدود به روش آموزش خاصّی نیستند؛ پس در تعمیم از دو نُرون به چند نُرون در لایهٔ میانی، می‌توان نتیجه گرفت که یک شبکهٔ عصبیِ پس‌انتشار سه لایه مانند شبکهٔ عصبیِ RBF گرانولی با آموزش پارامترها به روش ««الگوریتم خوشه‌بندی K-Means»» بتواند با تقریب بهتری نسبت به یک شبکهٔ عصبیِ RBF با همان تعداد نُرون در لایهٔ میانی رفتار نماید.

مراجع

[1] X. Wu, Y. Wang, W. Liu, Z. Zhu, and Y. Tan, "Application of Nonlinear Filtering Trained RBF Networks to Multi-step Prediction of Time-series with Delayed Observations," Journal of Information & Computational Science vol. 7: 3, pp. 385-393, 2011. Available at http://www.joics.com.

[2] M. Bahita, and K. Belarbi, "Neural Feedback Linearization Adaptive Control for Affine Nonlinear Systems Based on Neural Network Estimator," Serbian Journal of Electrical Engineering vol. 8, no. 3, pp. 307-323, November 2011.

[3] D. R. Brillinger, "Time Series: Data Analysis and Theory," vol. 36, Siam, 2001.

[4] Y. Yao, "Interval Sets and Interval-set Algebras," In Cognitive Informatics, ICCI'09, 8th IEEE International Conference on, pp. 307-314, 2009.

[5] X. Gao, "Some properties of continuous uncertain measure," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 17, no. 3, pp. 419-426, 2009.

[6] C.-F. Juang, R.-B. Huang and Y.-Y. Lin, "A recurrent self-evolving interval type-2 fuzzy neural network for dynamic system processing," Fuzzy Systems, IEEE Transactions on 17, no. 5, pp. 1092-1105, 2009.

[7] C.-F. Juang, and Y. W. Tsao, "A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning," IEEE Transactions on Fuzzy Systems, vol. 16, no. 6, pp. 1411-1424, December 2008.

[8] S. Guohua, L. Mingfeng, L. Teik C, "Enhanced Filtered-X Least Mean M-estimate Algorithm for Active Impulsive Noise Control," Applied Acoustics, 90: pp. 31-41, 2015.

[9] B. Santosh Kumar, D. Debi Prasad, S. Bidyadhar, "Functional Link Artificial Neural Network Applied to Active Noise Control of a Mixture from Tonal and Chaotic Noise," Applied Soft Computing, 23: pp. 51-60, 2014.

[10] M. Ahmadieh Khanesar, E. Kayacan, M. Teshnehlab, and O. Kaynak, "Analysis of the Noise Reduction Property of Type-2 Fuzzy Logic Systems Using a Novel Type-2 Membership Function," IEEE Transactions on Fuzzy Systems, Man, and Cybernetics-part B: Cybernetics, vol. 41, no. 5, October 2011.

[11] M. Ahmadieh Khanesar, M. Teshnehlab, E. Kayacan, and O. Kaynak, "A Novel Type-2 Fuzzy Membership Function: Application to the Prediction of Noisy Data," In Computational Intelligence for Measurement Systems and Applications (CIMSA), 2010 IEEE International Conference on, pp. 128-133. IEEE, 2010.

[12] B. Anja, et al, "Wavelet Based Noise Reduction in CT-images Using Correlation Analysis," Medical Imaging, IEEE Transactions on, 27.12, pp. 1685-1703, 2008.

[13] S. Stefan, et al, "Fuzzy Random Impulse Noise Reduction Method," Fuzzy Sets and Systems, 158.3, pp. 270-283, 2007.

[14] S. Stefan, et al, "A Fuzzy Impulse Noise Detection and Reduction Method," Image Processing, IEEE Transactions on, 15.5, pp. 1153-1162, 2006.

[15] T. Li; J. Jean, "Adaptive Volterra Filters for Active Control of Nonlinear Noise Processes," Signal Processing, IEEE Transactions on, 49.8, pp. 1667-1676, 2001.

[16] L. Chien-Cheng, Y.-C. Chiang, C.-Y. Shih and C.-L. Tsai, "Noisy Time-series Prediction Using M-estimator based Robust Radial Basis Function Neural Networks with Growing and Pruning Techniques," Expert Systems with Applications 36, no. 3, 4717-4724, 2009.

[17] Y. Y. Yao, "Human-inspired Granular Computing," Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation, pp. 1-15, 2010.

[18] Z. Meciarova, "Granular Computing and its Applications in RBF with Cloud Activation Function," Journal of Informational, Control and Management Systems, vol. 7, Issue 1, pp. 7-16, 2009.

[19] Y. Yao, and Z. Ning, "Granular Computing," Wiley Encyclopedia of Computer Science and Engineering (2008).

[20] S. Dick, A. Tappenden, C. Badke, and O. Olarewaju, "A Novel Granular Neural Network Architecture," In Fuzzy Information Processing Society, NAFIPS'07. Annual Meeting of the North American, pp. 42-47, IEEE, 2007.

[21] B. Andrzej, and W. Pedrycz, "The Roots of Granular Computing," In GrC, pp. 806-809, 2006.

[22] J. Abdi, B. Moshiri, and A. Khaki-Sedigh, "Comparison of RBF and MLP Neural Networks in Short-term Traffic Flow Forecasting," In Power, Control and Embedded Systems (ICPCES), 2010 International Conference on, pp. 1-4, IEEE, 2010.

[23] D. Marcek, M. Marcek, and J. Babel, "Granular RBF NN Approach and Statistical Methods Applied to Modeling and Forecasting High Frequency Data," International Journal of Computational Intelligence Systems, vol. 2, no. 4, pp. 353-364, December, 2009.

[24] G. B. Huang, P. Saratchandran, and N. Sundararajan, "A Generalized Growing and Pruning RBF (GGAP-RBF) Neural Network for Function Approximation," IEEE Transactions on Neural Networks, vol. 16, no. 1, pp. 57-67, January 2005.

[25] N. Jankowski, and V. Kadirkamanathan, "Statistical Control of Growing and Pruning in RBF-Like Neural Networks," In Third Conference on Neural Networks and Their Applications, pp. 663-670, 1997.

[26] J.-S. Roger Jang, and C.-T. Sun, "Functional Equivalence between Radial Basis Function Networks and Fuzzy Inference Systems," IEEE Transactions on Neural Networks, vol. 4, no. 1, pp. 156-159, January 1993.

[27] C. H. Skiadas, and C. Skiadas, "Chaotic Modeling and Simulation: Analysis of Chaotic Models, Attractors and Forms," Chapman and Hall/CRC Press is an imprint of Taylor and Francis Group, LLC, 2009, ISBN: 978-1-4200-7900-5 (hardcover: alk. paper).

[28] J. Principe and J.-M. Kou, "Dynamic Modeling of Chaotic Time-series with Neural Networks," Advances in Neural Information Processing Systems, pp. 311-318, 1995.

[29] V. Isham, "Statistical Aspects of Chaos, A Review," in: Networks and Chaos-Statistical and Probabilistic Aspects," pp. 124-200, Springer US, 1993.

[30] M. C. Mackey, and L. Glass, "Oscillation and Chaos in Physiological Control System," Sicence 197, no. 4300, pp. 287-289, 1977.

[31] B. Henon, "A Two Dimensional Mapping with a Strange Attractor," Communications in Mathematical Physics, vol. 50, issue 1, pp. 69-77, 1976.

[32] E. N. Lorenz, "Deterministic Nonperiodic Flow," Journal of the Atmospheric Sciences 20, no. 2, pp. 130-141, 1963.

[33] M. Teshnehlab, and K. Watanabe, "Intelligent Control Based on Flexible Neural Networks," Kluwer Academic publisher, 1999.

[34] A. Lapedes, and R. Farber, "Nonlinear Signal Processing Using Neural Networks," IEEE Conference on Neural Information Processing System, Natural and Synthetic. Proc., Mag., pp. 422, 1987.

[35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Back Propagation Errors," Nature, 323, pp. 533-536, 1986.

[36] V. Seydi Ghomsheh, M. Aliyari Shoorehdeli, and M. Teshnehlab, "Training ANFIS structure with modified PSO algorithm," 15th Mediterranean Conference on Control & Automation, July 27-29, 2007, Athens-Greece.

[37] D. Kukolj and E. Levi, "Identification of complex systems based on neural and Takagi-Sugeno fuzzy model," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 1, pp. 272–282, Feb. 2004

[38] V. Kreinovich, "From Interval and Probabilistic Granules to Granules of Higher Order," 2010.

[39] P. Billingsly, "Probability and Measure," John Wiley & Sons, 2008.

[40] S. Resnick, "Adventures in stochastic processes," Birkhäuser, Boston, 1992.

[41] A. Papoulis, and S. U. Pillai, "Probability, Random Variables and Stochastic Processes." Tata McGraw-Hill Education, 2002.

[۴۲] خاکی‌صدیق، علی (۱۳۸۳). ارزیابی روش‌های پیش‌بینی قیمت سهام و ارائهٔ مُدل بهینه. پژوهشکدهٔ پولی و بانکی بانک مرکزی جمهوری اسلامی ایران.

[۴۳] علیاری شوره دلی، مهدی، محمد تشنه لب و علی خاکی صدیق (۱۳۸۳). پیش‌بینی کوتاه مدت آلودگی هوا با کمک شبکه های عصبی پرسپترون چندلایه وخط حافظه دار تاخیر. ششمین کنفرانس سراسری سیستم های هوشمند، کرمان، دانشگاه شهید باهنر کرمان. http://www.civilica.com/Paper-ICS06-ICS06_025.html

[۴۴] ظهوری زنگنه، بختیار و عباس شولائی (۱۳۷۸). شبیه‌ساز(سیمولاتور) شبکه‌های HVDC/AC با گام بهینه. پانزدهمین کنفرانس بین المللی برق، تهران، شرکت توانیر، پژوهشگاه نیرو، http://www.civilica.com/Paper-PSC15-PSC15_067.html