

# یادگیری متریک بر اساس فاصله $\chi^2$ سریع برای دسته‌بندی داده‌های هیستوگرامی با دسته‌بندی کننده KNN

حمید صادقی<sup>۱</sup>، دانشجوی دکتری؛ ابوالقاسم اسدالله راعی<sup>۲</sup>، دانشیار

۱- دانشکده مهندسی برق - دانشگاه صنعتی امیرکبیر - تهران - ایران - hamid.sadeghi@aut.ac.ir

۲- دانشکده مهندسی برق - دانشگاه صنعتی امیرکبیر - تهران - ایران - raie@aut.ac.ir

**چکیده:** مقایسه داده‌ها یک مسأله بنیادی و پرکاربرد در یادگیری ماشین است. در دهه گذشته تحقیقات فراوانی در زمینه یادگیری متریک انجام شده است؛ از کاربردهای یادگیری متریک می‌توان به خوشه‌بندی و دسته‌بندی داده‌ها اشاره کرد. در این مقاله یک روش یادگیری متریک مناسب برای استفاده در مسائل بینایی ماشین ارائه می‌شود. اکثر ویژگی‌هایی که در بینایی ماشین استفاده می‌شوند، هیستوگرامی هستند؛ اما روش‌های یادگیری متریک اغلب بر مبنای فاصله ماهالانوبیس طراحی شده‌اند که در ویژگی‌های هیستوگرامی کارایی مناسبی ندارد. در این تحقیق یک روش جدید یادگیری متریک برای داده‌های هیستوگرامی بر مبنای فاصله مربع کای ( $\chi^2$ ) اصلاح شده ارائه می‌شود. فاصله  $\chi^2$  در دسته‌بندی داده‌های هیستوگرامی دارای دقت بالاتری نسبت به فاصله اقلیدسی است، اما هزینه محاسباتی آن نیز بالاتر است. در این مقاله یک رابطه تقریبی برای فاصله  $\chi^2$  پیشنهاد می‌شود و بخشی از محاسبات را به مرحله استخراج ویژگی (که به صورت غیربرخط<sup>۱</sup> قابل محاسبه است) منتقل می‌کند؛ به این ترتیب سرعت مقایسه ویژگی‌ها افزایش می‌یابد. آزمایش‌ها بر روی پایگاه‌های داده مختلف نشان می‌دهد که روش یادگیری متریک پیشنهادی دارای دقت بالایی در دسته‌بندی داده‌های هیستوگرامی مختلف نسبت به روش‌های موجود است. همچنین معیار تقریبی برای فاصله  $\chi^2$ ، با حفظ دقت، سرعت مقایسه داده‌ها را ۲/۵ برابر افزایش می‌دهد.

**واژه‌های کلیدی:** یادگیری متریک، فاصله مربع کای سریع، دسته‌بندی هیستوگرام، دسته‌بندی کننده KNN.

## Metric Learning based on Fast $\chi^2$ Distance for Histogram Data Classification via KNN Classifier

Hamid Sadeghi<sup>1</sup>, PhD Student; Abolghasem-A. Raie<sup>2</sup>, Associate Professor

1- Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran, Email: hamid.sadeghi@aut.ac.ir

2- Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran, Email: raie@aut.ac.ir

**Abstract:** Data comparison is a fundamental problem in machine learning research. Since, metric learning has various applications in clustering and classification problems, it is attracted much attention in the last decades. In this paper, an appropriate metric learning method is presented to utilize in machine vision problems. Common features in machine vision are often histogram; however, metric learning methods are usually designed based on Mahalanobis distance which is not applicable in histogram features. In this study, a new metric learning method based on modified chi-squared distance ( $\chi^2$ ) for histogram data is presented. In histogram data classification,  $\chi^2$  distance is more accurate than Euclidean one; however, its computational cost is higher than Euclidean distance. In this paper, a  $\chi^2$  distance approximated formulation where a part of its computations is moved into the feature extraction step in offline phase is proposed. Consequently, computational cost of feature comparison is reduced. Experiments on different datasets show that the proposed metric learning method is more accurate than the existing ones in histogram data classification. Moreover, the approximated  $\chi^2$  distance increases feature comparison speed about 2.5 times without loss of accuracy.

**Keywords:** Metric learning, fast chi-squared distance, histogram classification, KNN classifier.

تاریخ ارسال مقاله: ۱۳۹۶/۶/۷

تاریخ اصلاح مقاله: ۱۳۹۷/۲/۱۳

تاریخ پذیرش مقاله: ۱۳۹۷/۳/۱۴

نام نویسنده مسئول: ابوالقاسم اسدالله راعی

نشانی نویسنده مسئول: ایران - تهران - خیابان حافظ - دانشگاه صنعتی امیرکبیر - دانشکده مهندسی برق.

## ۱- مقدمه

مقایسه داده‌ها یک مسأله بنیادی در تحقیقات یادگیری ماشین است؛ به طوری که تشخیص صحیح مشابه یا نامشابه بودن داده‌ها می‌تواند راه‌گشای مسائل مختلف در یادگیری ماشین باشد. به عنوان مثال، استفاده از یک معیار مقایسه مناسب، در مسائلی که تنها یک داده تعلیم در هر دسته<sup>۲</sup> وجود دارد (مانند شناسایی چهره بر اساس یک تصویر) و نیز مسائلی که داده‌های تعلیم و تابع هدف بسیار پیچیده دارند، می‌تواند راه‌حلی کاملاً منطقی باشد. با توجه به کاربردهای مختلف در مسائلی همچون خوشه‌بندی<sup>۳</sup> و دسته‌بندی<sup>۴</sup> داده‌ها، در دهه گذشته تحقیقات بسیاری در زمینه یادگیری متریک انجام شده است [۱، ۲]. از کاربردهای عملی یادگیری متریک می‌توان به علامت‌گذاری خودکار تصویر<sup>۵</sup> [۳]، شناسایی چهره [۴] و تشخیص خویشاوندی [۵]، [۶] در تصویر اشاره نمود. در این مقاله راهکار جدیدی برای یادگیری متریک در داده‌های هیستوگرامی که کاربرد بسیار زیادی در بینایی ماشین دارند، ارائه می‌شود.

هدف اصلی یادگیری متریک، بهبود معیار مقایسه داده‌ها بر اساس یک مجموعه آموزشی است [۲]. بسیاری از روش‌های یادگیری متریک، سعی در بهبود معیار فاصله ماهالانویس<sup>۶</sup> دارند. فاصله ماهالانویس بین دو بردار  $D$  بعدی  $\vec{x}_i$  و  $\vec{x}_j$  به صورت رابطه (۱) تعریف می‌شود.

$$D_A(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^T A (\vec{x}_i - \vec{x}_j) \quad (1)$$

که در این رابطه،  $A$  یک ماتریس مربعی با ابعاد  $D \times D$  است که با یادگیری بر روی مجموعه آموزشی به گونه‌ای به دست می‌آید که قابلیت تفکیک‌پذیری دسته‌های مختلف را برای معیار ماهالانویس بهبود دهد. اولین تلاش انجام شده در زمینه یادگیری فاصله ماهالانویس، بر اساس یک فرمول‌نویسی محذب و بدون تنظیم‌کننده بوده است [۷]. پس از آن تلاش‌های دیگری در این زمینه انجام شده و تاکنون نیز ادامه داشته است [۸-۱۷].

در [۸] روشی بر اساس رابطه  $A = M^T W M$  ارائه شده است که در آن ماتریس  $M$  ثابت و معلوم و  $W$  یک ماتریس قطری است. در این روش، فرآیند یادگیری بر روی ماتریس قطری  $W$  انجام می‌شود. با تعریف انجام شده در این روش، ماتریس  $A$  مثبت نیمه معین<sup>۷</sup> (PSD) خواهد بود و نیازی به اعمال قید PSD بودن ماتریس یادگیری و تصویرسازی این ماتریس بر روی مخروط PSD نیست. علاوه بر لزوم تعیین دستی ماتریس  $M$ ، نقطه ضعف اصلی این روش قابلیت تعمیم‌پذیری کمتر آن نسبت به یادگیری فاصله ماهالانویس متداول است؛ چراکه در این روش، ماتریس  $W$  قطری بوده و در فرآیند یادگیری تنها یک وزن‌دهی ویژگی‌ها آموزش داده می‌شود و ارتباط و اثر متقابل ویژگی‌ها با یکدیگر در نظر گرفته نمی‌شود.

روش ITML<sup>۸</sup> [۹] یکی از مهم‌ترین تحقیقات در زمینه یادگیری فاصله ماهالانویس است که برای نخستین بار تنظیم‌کننده واگرایی<sup>۹</sup> LogDeg را معرفی نموده است. این تنظیم‌کننده در بسیاری از تحقیقات دیگر مورد استفاده قرار گرفته است. یکی از محدودیت‌های

این روش، لزوم مقداردهی دستی اولیه در یکی از ماتریس‌های یادگیری است که می‌تواند کیفیت یادگیری را تحت تأثیر خود قرار دهد. استفاده از واگرایی Jeffrey یکی دیگر از روش‌های بهینه‌سازی در یادگیری فاصله ماهالانویس بوده است [۱۳]. در [۱۶] یادگیری متریک ماهالانویس به صورت یک مسأله دسته‌بندی هسته<sup>۱۰</sup> تعریف شده است و با کمک این تعریف، از یادگیری تکراری ماشین بردار پشتیبان<sup>۱۱</sup> برای حل مسأله یادگیری متریک استفاده شده است.

یکی دیگر از تحقیقات انجام شده در زمینه یادگیری متریک که مورد استفاده فراوان نیز قرار گرفته است، روش LMNN<sup>۱۲</sup> است [۱۴]. این روش برای دسته‌بندی‌کننده  $K$ -نزدیک‌ترین همسایگی<sup>۱۳</sup> (KNN) و بر اساس فاصله ماهالانویس طراحی شده است. دسته‌بندی‌کننده KNN یک روش پرکاربرد در مسائل مختلف بازنمایی الگو و دسته‌بندی داده‌های متنوع است [۱۸، ۱۹]. در [۱۵] یک روش بهینه‌سازی مبتنی بر مقادیر ویژه برای بهبود LMNN ارائه شده است؛ این روش نیز بر اساس فاصله ماهالانویس طراحی شده است. در سال‌های بعد، روش LMNN برای استفاده در داده‌های هیستوگرامی بر اساس فاصله مربع کای<sup>۱۴</sup> ( $\chi^2$ ) توسعه داده شد [۲۰]. می‌توان گفت این روش، معروف‌ترین الگوریتم یادگیری متریک برای فاصله  $\chi^2$  است؛ اما از نقاط ضعف آن می‌توان به عدم استفاده از تنظیم‌کننده<sup>۱۵</sup> در فرآیند آموزش و وقوع بیش‌تعلیم<sup>۱۶</sup> بر روی داده‌های آموزشی و همچنین محذب نبودن تابع هزینه تعریف شده اشاره کرد. در یک تحقیق دیگر در سال ۲۰۱۵ [۳۸]، یک روش یادگیری متریک  $\chi^2$  برای دسته‌بندی داده‌های هیستوگرامی پیشنهاد شده است. در این روش، اصلاح رابطه  $\chi^2$  برای یادگیری متریک مانند روش ارائه شده در [۲۰] بوده است. از نقاط ضعف این روش می‌توان به نیاز به افکنش ماتریس یادگیری به فضای غیرمربک<sup>۱۷</sup> در هر تکرار از الگوریتم یادگیری (به منظور حفظ خاصیت هیستوگرامی داده‌ها) اشاره نمود. نویسندگان برای کاهش این هزینه محاسباتی، از یک مسأله بهینه‌سازی بدون قید استفاده کرده‌اند که در این حالت نیز استفاده از تنظیم‌کننده و تعمیم‌پذیری الگوریتم بسیار مشکل شده است.

همان‌طور که در رابطه (۱) نیز مشهود است، معیار فاصله ماهالانویس متناسب با فضاهای متعامد تعریف شده است. درحالی‌که اغلب ویژگی‌هایی که در مسائل بینایی ماشین و پردازش تصویر مورد استفاده قرار می‌گیرند، توصیف هیستوگرامی بافت تصویر هستند و استفاده از معیار فاصله ماهالانویس در این فضای غیرمتعامد هیستوگرامی کارایی چندانی ندارد. نمونه‌ای از این ویژگی‌های هیستوگرامی، ویژگی HOG<sup>۱۸</sup> است که در [۲۱] برای تخمین سن بر اساس تصاویر چهره مورد استفاده قرار گرفته است. فاصله  $\chi^2$  شناخته‌شده‌ترین معیار مقایسه داده‌های هیستوگرامی است که با موفقیت در کاربردهای مختلف همچون شناسایی چهره مورد استفاده قرار گرفته است [۲۲، ۲۳]. در این مقاله روش جدیدی برای یادگیری متریک بر اساس معیار  $\chi^2$  پیشنهاد می‌شود. برای این کار، رابطه ریاضی معیار  $\chi^2$  به نحو مناسبی برای یادگیری متریک اصلاح می‌شود. همچون

هیستوگرامی شناخته شده در یادگیری متریک نسبت به سایر روش‌های موجود دارای دقت بالاتری است.

- یکی از راه‌های پیشگیری از بیش‌تعلیم در یادگیری متریک، استفاده از تنظیم‌کننده است. در این مطالعه، از تکنیک حذف تصادفی به‌عنوان تنظیم‌کننده در یادگیری متریک استفاده شده است و به‌صورت ریاضی نشان داده می‌شود که روش پیشنهادی معادل با استفاده از یک رابطه تنظیم‌کننده نسبتاً پیچیده در حین یادگیری است.

- نقطه ضعف معیار فاصله  $\chi^2$  نسبت به معیار اقلیدسی، هزینه محاسباتی بیشتر آن است. در این مقاله یک رابطه تقریبی برای معیار فاصله  $\chi^2$  ارائه می‌شود که با حفظ دقت این معیار، سرعت محاسبات آن را  $2/5$  برابر می‌کند.

ادامه مطالب موجود در این مقاله بدین شرح است: در بخش بعدی، روش پیشنهادی یادگیری متریک و همچنین رابطه معیار فاصله  $\chi^2$  تقریبی و تلفیق این دو شرح داده می‌شود. در بخش سوم به شرح پایگاه‌های داده و آزمایش‌های انجام شده بر روی آنها و ارائه و تحلیل نتایج پرداخته می‌شود. نهایتاً بخش چهارم مربوط به جمع‌بندی این مقاله است.

## ۲- روش پیشنهادی

در این بخش، ابتدا روش یادگیری متریک LMNN به‌صورت مختصر معرفی می‌شود؛ در ادامه فاصله  $\chi^2$  و رابطه پیشنهادی مبتنی بر آن برای یادگیری متریک بیان می‌شود. سپس به شرح تابع هزینه پیشنهادی و تکنیک حذف تصادفی به‌عنوان تنظیم‌کننده آن و نیز الگوریتم یادگیری این تابع پرداخته می‌شود. در انتها رابطه  $\chi^2$  تقریبی پیشنهادی بیان و با الگوریتم یادگیری متریک تلفیق می‌شود.

### ۲-۱ یادگیری متریک LMNN

روش یادگیری متریک LMNN به‌منظور یادگیری فاصله ماهالانوبیس در دسته‌بندی کننده KNN ارائه شده است [۱۴]. در این روش، یادگیری با این هدف انجام می‌شود که: (الف) K همسایه هر داده دارای برچسب مشابه با این داده باشند و (ب) داده‌های نامشابه با یک حاشیه بزرگ نسبت به این داده، قابل تفکیک باشند. فرض کنید که  $X = \{(\vec{x}_i, y_i) | i = 1, 2, \dots, N\}$  مجموعه آموزشی متشکل از N داده باشد که در آن  $\vec{x}_i$  نشان‌دهنده داده i ام با برچسب  $y_i$  از فضای D بعدی ( $\vec{x}_i \in \mathbb{R}^D$ ) باشد؛ فاصله ماهالانوبیس بین دو بردار  $\vec{x}_i$  و  $\vec{x}_j$  در رابطه (۱) نشان داده شد. با فرض اینکه  $A = L^T L$  باشد، می‌توان رابطه (۱) را به‌صورت رابطه (۲) بازنویسی کرد.

$$D_A(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^T L^T L (\vec{x}_i - \vec{x}_j) = \left( L(\vec{x}_i - \vec{x}_j) \right)^T \left( L(\vec{x}_i - \vec{x}_j) \right) = \|L(\vec{x}_i - \vec{x}_j)\|^2 \quad (2)$$

معیار فاصله ماهالانوبیس (رابطه (۱))، رابطه پیشنهادی جدید دارای یک ماتریس یادگیری است که با یادگیری بر روی مجموعه آموزشی به‌دست خواهد آمد.

برای یادگیری ماتریس موجود در معیار فاصله مبتنی بر  $\chi^2$  پیشنهادی، یک تابع هزینه محذب ارائه شده است. این تابع هزینه بر اساس بهبود دسته‌بندی کننده KNN طراحی شده است. در ادامه تابع هزینه ارائه شده با استفاده از یک الگوریتم مبتنی بر کاهش گرادیان تصادفی<sup>۱۹</sup> (SGD) بهینه‌سازی شده است.

یکی از چالش‌های موجود در یادگیری متریک، تعداد بالای پارامترهای یادگیری است؛ این مسأله به همراه تعداد کم داده آموزشی باعث به‌وجود آمدن بیش‌تعلیم بر روی مجموعه آموزشی می‌شود. یکی از راه‌های پیشگیری از بیش‌تعلیم در یادگیری متریک، استفاده از تنظیم‌کننده است. در این مقاله از یک تکنیک حذف تصادفی<sup>۲۰</sup> جدید به‌عنوان تنظیم‌کننده برای پیشگیری از بیش‌تعلیم استفاده شده است. روش حذف تصادفی در اصل برای مقابله با بیش‌تعلیم در آموزش شبکه‌های عصبی عمیق ارائه شده است [۲۴]؛ در این روش، در حین آموزش شبکه عصبی با استفاده از هر نمونه آموزشی، نیمی از آشکارسازهای ویژگی‌ها به‌صورت تصادفی از فرآیند به‌روزرسانی کنار گذاشته شده و مابقی وزن‌ها به‌روزرسانی می‌شوند. روش حذف تصادفی در رگرسیون [۲۵، ۲۶] و برای نخستین بار در [۲۷] برای مسأله یادگیری متریک نیز مورد استفاده قرار گرفته است. در این مقاله نیز یک رابطه حذف تصادفی جدید برای یادگیری متریک  $\chi^2$  پیشنهاد شده است و نشان داده می‌شود که این روش از نظر ریاضی معادل با استفاده از یک تنظیم‌کننده نسبتاً پیچیده در حین یادگیری است.

همان‌طور که ذکر شد، معیار فاصله  $\chi^2$  در ویژگی‌های هیستوگرامی دارای دقت بالاتری نسبت به معیار فاصله ماهالانوبیس و اقلیدسی است؛ اما نقطه ضعف آن نسبت به فاصله اقلیدسی، محاسبات ریاضی بیشتر است. برای غلبه بر این مشکل، در این مقاله یک رابطه تقریبی از معیار فاصله  $\chi^2$  پیشنهاد می‌شود که مخرج این رابطه را به مرحله محاسبه بردار ویژگی منتقل می‌کند. به این ترتیب، زمان مربوط به محاسبه فاصله (تطبیق ویژگی) کاهش می‌یابد. معیار پیشنهادی بر روی چند پایگاه داده در مسائل شناسایی چهره و حالت چهره آزمایش شده است. نتایج به‌دست آمده نشان می‌دهد که دقت محاسبه فاصله در ویژگی‌های هیستوگرامی با استفاده از روش پیشنهادی تقریباً برابر با معیار  $\chi^2$  و سرعت محاسبه آن برابر با معیار فاصله اقلیدسی است. در ادامه این معیار پیشنهادی با ماتریس حاصل از یادگیری متریک پیشنهادی تلفیق شده است.

به‌طور کلی، نوآوری‌های اصلی این مقاله به شرح زیر است:

- در این مقاله، یک روش جدید یادگیری متریک برای معیار فاصله  $\chi^2$  ارائه شده است. هدف اصلی این روش برای استفاده در مقایسه داده‌های هیستوگرامی است. روش پیشنهادی بر روی داده‌های

$$\chi^2(\bar{x}_i, \bar{x}_j) = \sum_{b=1}^D \frac{(\bar{x}_i)_b - (\bar{x}_j)_b}{(\bar{x}_i)_b + (\bar{x}_j)_b} \quad (4)$$

که در آن،  $(\bar{x}_i)_b$  نشان دهنده عنصر  $b$  ام از هیستوگرام  $\bar{x}_i$  است. رابطه (۴) می‌تواند به صورت ماتریسی به شکل رابطه (۵) بازنویسی شود.

$$\begin{aligned} \chi^2(\bar{x}_i, \bar{x}_j) &= (\bar{x}_i - \bar{x}_j)^T (\text{diag}(\bar{x}_i + \bar{x}_j))^{-1} (\bar{x}_i - \bar{x}_j) \\ &= \left( (\bar{x}_i - \bar{x}_j) \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}} \right)^T \\ &\quad \left( (\bar{x}_i - \bar{x}_j) \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}} \right) \end{aligned} \quad (5)$$

که در این رابطه،  $\sqrt{\cdot}$  جذر عناصر ماتریس را محاسبه می‌کند و  $\text{diag}(V)$  عناصر بردار  $D$  بعدی  $V$  را در قطر اصلی یک ماتریس  $D \times D$  که عناصر غیر قطر اصلی آن صفر است، قرار می‌دهد. در این مقاله، رابطه (۶) برای یادگیری متریک در فاصله  $\chi^2$  پیشنهاد شده است.

$$\begin{aligned} \chi_A^2(\bar{x}_i, \bar{x}_j) &= \left( (\bar{x}_i - \bar{x}_j) \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}} \right)^T \\ &\quad A \left( (\bar{x}_i - \bar{x}_j) \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}} \right) \end{aligned} \quad (6)$$

که  $A$  ماتریس یادگیری با ابعاد  $D \times D$  است و سایر متغیرها مانند رابطه (۵) تعریف می‌شود. اگر بردارهای  $\bar{x}_{io}$  و  $\bar{x}_{joi}$  به ترتیب به صورت  $\bar{x}_i \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}}$  و  $\bar{x}_j \sqrt{(\text{diag}(\bar{x}_i + \bar{x}_j))^{-1}}$  تعریف شوند، می‌توان رابطه (۶) را به صورت رابطه (۷) ساده‌نویسی نمود.

$$\chi_A^2(\bar{x}_i, \bar{x}_j) = (\bar{x}_{ioj} - \bar{x}_{joi})^T A (\bar{x}_{ioj} - \bar{x}_{joi}) \quad (7)$$

با فرض  $A = L^T L$ ، رابطه (۷) به شکل رابطه (۸) قابل بازنویسی است.

$$\chi_A^2(\bar{x}_i, \bar{x}_j) = \chi_L^2(\bar{x}_i, \bar{x}_j) = \|L(\bar{x}_{ioj} - \bar{x}_{joi})\|^2 \quad (8)$$

همان‌طور که ملاحظه می‌شود، روابط به دست آمده مشابه با روابط بیان شده برای یادگیری متریک در فاصله ماله‌انوبیس هستند. به عبارت دیگر، یادگیری متریک بر اساس معیار  $\chi^2$  با یک تغییر متغیر به یک مسأله یادگیری متریک بر اساس فاصله ماله‌انوبیس تبدیل شده است.

## ۴-۲- تابع هزینه

بر اساس تابع هزینه‌ای که برای روش یادگیری متریک LMNN در رابطه (۳) بیان شد، تابع هزینه پیشنهادی برای یادگیری متریک  $\chi^2$  (رابطه (۸)) به صورت رابطه (۹) تعریف شده است.

$$\begin{aligned} \ell(L, X) &= (1 - \mu) \sum_{i,j=i} \|L(\bar{x}_{ioj} - \bar{x}_{joi})\|^2 + \\ &\quad \mu \sum_{i,j=i, i: y_i \neq y_j} [1 + \|L(\bar{x}_{ioj} - \bar{x}_{joi})\|^2 - \|L(\bar{x}_{ioi} - \bar{x}_{joi})\|^2]_+ \end{aligned} \quad (9)$$

که در این رابطه،  $L$  ماتریسی است که با یادگیری بر روی مجموعه آموزشی به دست می‌آید.

در روش LMNN، برای داده آموزشی  $\bar{x}_i$ ، تعداد  $K$  نزدیک‌ترین همسایه با برچسب مشابه، به عنوان داده‌های هدف<sup>۲۱</sup> در نظر گرفته می‌شود [۱۴]. نزدیک‌ترین همسایه‌ها با محاسبه فاصله اقلیدسی تعیین می‌شوند. در این مقاله، برای نمایش اینکه داده  $\bar{x}_j$  جزء داده‌های هدف  $\bar{x}_i$  است، از نشانه‌گذاری  $j \sim i$  استفاده شده است. برای انجام یک دسته‌بندی KNN موفق، لازم است که در بین  $K$  نزدیک‌ترین همسایه داده  $\bar{x}_i$ ، داده‌ای با برچسب نامشابه وجود نداشته باشد. بنابراین در روش LMNN، برای داده  $\bar{x}_i$  یک محیط در نظر گرفته می‌شود که دربردارنده داده‌های هدف به همراه یک حاشیه امنیت است. با این تعریف، داده‌های نامشابه‌ای که در این محیط قرار گرفته و دسته‌بندی را با مشکل روبرو می‌کنند، به عنوان داده‌های فریبکار<sup>۲۲</sup> در نظر گرفته شده‌اند. در یادگیری LMNN، باید تعداد داده‌های فریبکار پس از یادگیری، حداقل شود. به طور کلی، تابع هزینه معرفی شده برای یادگیری در این روش به صورت رابطه (۳) است.

$$\begin{aligned} \ell(L, X) &= (1 - \mu) \sum_{i,j=i} \|L(\bar{x}_i - \bar{x}_j)\|^2 + \\ &\quad \mu \sum_{i,j=i, i: y_i \neq y_j} [1 + \|L(\bar{x}_i - \bar{x}_j)\|^2 - \|L(\bar{x}_i - \bar{x}_i)\|^2]_+ \end{aligned} \quad (3)$$

که در این رابطه،  $\bar{x}$  و  $y$  به ترتیب داده آموزشی و برچسب آن،  $L$  ماتریس یادگیری (مانند رابطه (۲))،  $[u]_+ = \max(u, 0)$  تابع هزینه لولایی<sup>۲۳</sup> و  $\mu$  ضریب ثابت تعیین کننده تأثیر عبارت اول و دوم تابع هزینه است که در [۱۴] برابر با ۰.۵ قرار داده شده است. عبارت اول رابطه (۳) فواصل بزرگ بین داده  $\bar{x}_i$  و داده‌های هدف آن را جریمه می‌کند و بخش دوم این رابطه، داده‌های فریبکار را برای  $\bar{x}_i$  جریمه می‌کند. همان‌طور که ذکر شد، داده‌های فریبکار داده‌های نامشابه‌ای هستند که فاصله آنها تا  $\bar{x}_i$ ، از فاصله  $\bar{x}_i$  تا داده‌های هدف آن (به علاوه حاشیه امنیت ۱) کمتر باشد. استفاده از تابع هزینه لولایی باعث می‌شود که داده‌های نامشابه‌ای که در فاصله‌ای دور از  $\bar{x}_i$  قرار دارند، در تابع هزینه کلی وارد نشده و در نتیجه در فرآیند آموزش بی‌تأثیر باشند. به طور کلی، در فرآیند بهینه‌سازی LMNN، گرادینان عبارت اول در تابع هزینه، داده‌های هدف را به سمت داده مورد نظر نزدیک می‌کند و گرادینان بخش دوم، داده‌های فریبکار را از محیط پیرامون  $\bar{x}_i$  دور می‌کند.

## ۴-۲- اصلاح فاصله $\chi^2$ برای یادگیری متریک

معیار فاصله  $\chi^2$  مورد توجه بسیاری از محققین در مقایسه داده‌های هیستوگرامی قرار گرفته است [۲۳، ۲۲]. به عنوان مثال در [۲۲] از این معیار برای مقایسه ویژگی‌های هیستوگرامی استخراج شده از تصاویر چهره استفاده شده است. معیار فاصله  $\chi^2$  برای دو بردار (هیستوگرامی)  $\bar{x}_i$  و  $\bar{x}_j$  به صورت رابطه (۴) تعریف می‌شود.

روش حذف تصادفی که در اصل برای تعلیم شبکه‌های عصبی عمیق معرفی شده است، برای یادگیری متریک نیز مورد استفاده قرار گرفته است [۲۷]. این روش برای جلوگیری از وقوع بیش‌تعلیم الگوریتم بر روی داده‌های آموزشی انجام می‌شود. روش حذف تصادفی برای یادگیری متریک می‌تواند قبل از مرحله ۵ در الگوریتم ۱ اعمال شود. برای این کار، طبق رابطه (۱۲)، عناصر ماتریس یادگیری در متغیرهای تصادفی برنولی ضرب می‌شوند.

$$\hat{A}_{ij} = \delta_{ij} A_{ij}, \quad i, j = 1, 2, \dots, d \quad (12)$$

که  $\delta_{ij}$  یک متغیر تصادفی برنولی با شرط  $\delta_{ij} = \delta_{ji}$  است. این شرط، خاصیت متقارن بودن ماتریس  $\hat{A}_{ij}$  را حفظ می‌کند. در این حالت، به‌روزرسانی ماتریس  $A$  در مرحله ۵ الگوریتم SGD به‌صورت رابطه (۱۳) تعریف می‌شود.

$$A_t = \hat{A}_{t-1} - \eta \nabla \ell \quad (13)$$

که در آن،  $\nabla \ell$  گرادیان تابع هزینه  $\ell$  و گام یادگیری است.

#### ۲ ۵ حذف تصادفی به‌عنوان تنظیم‌کننده نرّم فروبنیوس<sup>۲۶</sup>

نرّم فروبنیوس یکی از تنظیم‌کننده‌های پرکاربرد در یادگیری متریک است [۱، ۲، ۲۷، ۲۹، ۳۰] که برای جلوگیری از بیش‌تعلیم بر روی داده‌های آموزشی مورد استفاده قرار می‌گیرد. شکل استاندارد یک مسأله یادگیری متریک با تنظیم‌کننده نرّم فروبنیوس به‌صورت رابطه (۱۴) تعریف می‌شود.

$$\min_{\text{constraints}} \{ \ell_F(A, X) = \ell(A, X) + \frac{\beta}{2\eta} \|A\|_F^2 \} \quad (14)$$

که در این رابطه،  $X$  مجموعه آموزشی،  $A$  ماتریس یادگیری،  $\|\cdot\|_F$  عملگر نرّم فروبنیوس،  $\eta$  گام تصحیح گرادیان در فرآیند یادگیری و  $\beta$  ضریب تأثیر تنظیم‌کننده است. در حل مسأله بهینه‌سازی نشان داده شده در رابطه (۱۴) با استفاده از روش کاهش گرادیان، به‌روزرسانی ماتریس  $A$  به‌صورت رابطه (۱۵) تعریف می‌شود.

$$A_t = A_{t-1} - \eta \nabla \ell_F \quad (15)$$

با توجه به اینکه  $\nabla \|A\|_F^2 = 2A$  است خواهیم داشت:  $\nabla \ell_F = \frac{\beta}{\eta} A + \nabla \ell$  و می‌توان رابطه (۱۵) را به‌صورت رابطه (۱۶) بازنویسی کرد.

#### الگوریتم ۱: یادگیری متریک به روش SGD

- ۱: ورودی‌ها: پایگاه داده آموزشی  $X \in \mathbb{R}^{D \times N}$  و گام یادگیری  $\eta$
- ۲: مقداردهی اولیه  $A_0$  و  $T$
- ۳: حلقه تکرار  $t = 1$  تا  $T$
- ۴: انتخاب نمونه تصادفی  $\{\bar{x}_i, \bar{x}_j, \bar{x}_i\}_{j \sim i}$
- ۵: به‌روزرسانی  $A_t = A_{t-1} - \eta \nabla \ell$
- ۶:  $\prod_{psd} (A_t)$
- ۷: اتمام حلقه تکرار
- ۸: خروجی:  $\bar{A}$

که در این رابطه، بردارهای  $\bar{x}_{ioj}$  و  $\bar{x}_{joi}$  مانند رابطه (۷) و (۸) و سایر متغیرها مانند رابطه (۳) هستند.

تابع هزینه ارائه شده در رابطه (۹)، در عناصر ماتریس تبدیل خطی  $L$  محدب نیست [۱۴]. برای کمینه‌سازی این تابع می‌توان از روش کاهش گرادیان استفاده نمود؛ اما نتیجه نهایی این روش، به‌شدت وابسته به مقادیر اولیه ماتریس  $L$  بوده و به کمینه‌های موضعی می‌رسد. برای غلبه بر این مشکل، رابطه (۹) می‌تواند به‌صورت یک مسأله برنامه‌نویسی نیمه‌معین<sup>۲۴</sup> (SDP) [۲۸] بازنویسی شود. در یک مسأله SDP، بهینه‌سازی با قید مثبت و نیمه معین بودن ماتریس انجام می‌شود. برای این کار، بر اساس تعریف  $A = L^T L$  و رابطه (۷)، بهینه‌سازی به شکل رابطه (۱۰) و با قید نیمه‌معین مثبت بودن ماتریس  $A (A \succeq 0)$  انجام شده است. رابطه (۱۰) نشان دهنده یک مسأله بهینه‌سازی محدب بوده و گرادیان آن قابل محاسبه است (پیوست الف).

$$\ell(A, X) = (1 - \mu) \sum_{i, j \sim i} \chi_A^2(\bar{x}_i, \bar{x}_j) + \mu \sum_{i, j \sim i; y_i \neq y_j} [1 + \chi_A^2(\bar{x}_i, \bar{x}_j) - \chi_A^2(\bar{x}_i, \bar{x}_j)]_+, \quad A \succeq 0 \quad (10)$$

که  $\chi_A^2$  طبق رابطه (۷) تعریف شده است.

برای نوشتن رابطه (۱۰) به‌صورت یک مسأله SDP، باید آن را به یک شکل استاندارد تبدیل کرد. این کار می‌تواند با تعریف متغیرهای کمکی<sup>۲۵</sup> که رفتاری مشابه با تابع اتلاف لولایی دارند، انجام شود. برای این منظور، متغیر کمکی به شکل  $\xi_{ijl}$  برای داده  $\bar{x}_i$  و داده هدف  $j \sim i$  و داده فریبکار  $\bar{x}_i$  تعریف شده است [۱۴]. این متغیر نشان می‌دهد که داده فریبکار  $\bar{x}_i$  تا چه اندازه وارد محیط پیرامون داده  $\bar{x}_i$  با داده هدف  $j \sim i$  و حاشیه امنیت آن شده است. با این تعریف، مسأله بهینه‌سازی برای یادگیری متریک  $\chi^2$  به‌صورت رابطه (۱۱) تعریف شده است.

$$\min (1 - \mu) \sum_{i, j \sim i} (\bar{x}_{ioj} - \bar{x}_{joi})^T A (\bar{x}_{ioj} - \bar{x}_{joi}) + \mu \sum_{i, j \sim i; y_i \neq y_j} \xi_{ijl} \begin{cases} (\bar{x}_{ioi} - \bar{x}_{joi})^T A (\bar{x}_{ioi} - \bar{x}_{joi}) - (\bar{x}_{ioj} - \bar{x}_{joi})^T A (\bar{x}_{ioj} - \bar{x}_{joi}) \geq 1 - \xi_{ijl} \\ \xi_{ijl} \geq 0 \\ A \succeq 0 \end{cases} \quad (11)$$

در این رابطه،  $A$  ماتریس یادگیری،  $\xi_{ijl}$  متغیر کمکی،  $\mu$  ضریب ثابت تأثیر عبارت اول و دوم و  $\bar{x}_{ioj}$  و  $\bar{x}_{joi}$  مانند رابطه (۷) و (۸) هستند.

#### ۲ ۴ تنظیم‌کننده تابع هزینه با روش حذف تصادفی

در بخش‌های پیش، یک مسأله بهینه‌سازی برای متریک  $\chi^2$  شرح داده شد. الگوریتم SGD یکی از روش‌های پرکاربرد در حل مسائل بهینه‌سازی در یادگیری متریک است [۱، ۹، ۱۱، ۲۷، ۲۹، ۳۰]. این روش به‌صورت کلی در الگوریتم ۱ توصیف شده است.

با توجه به رابطه (۱۹) و رابطه (۱۲) می توان نوشت:

$$E[\hat{A}_{t-1}] = \begin{cases} A_{i,j}^{t-1}(1-2\beta) & i = j \\ A_{i,j}^{t-1}(1-\beta) & i \neq j \end{cases} \quad (20)$$

استفاده از روش حذف تصادفی معرفی شده در رابطه (۱۹)، معادل با حل بهینه سازی نشان داده شده در رابطه (۲۱) است.

$$\min_{\text{constraints}} \ell(A, X) + \frac{\beta}{2\eta} (\|A\|_F^2 + \sum_{i=1}^D A_{i,j}^2) \quad (21)$$

همان طور که در رابطه (۲۱) ملاحظه می شود، تنظیم کننده استفاده شده، معادل با نرم فروبنیوس ماتریس A به علاوه نرم ۱ قطر اصلی این ماتریس است. به عبارت دیگر، علاوه بر جریمه شدن نرم فروبنیوس ماتریس A، نرم ۱ قطر اصلی نیز جریمه می شود. بنابراین، قطر اصلی ماتریس A، یکبار در تنظیم کننده نرم فروبنیوس و یکبار دیگر در تنظیم کننده نرم ۱ شرکت داشته است، اما سایر عناصر این ماتریس فقط در تنظیم کننده نرم فروبنیوس حضور داشته اند. مسأله بهینه سازی رابطه (۲۱) به ظاهر پیچیده است، اما با استفاده از روش حذف تصادفی معرفی شده در الگوریتم ۲، به راحتی قابل حل است.

علاوه بر استفاده از پارامتر حذف تصادفی، الگوریتم ۲ نسبت به الگوریتم ۱ در اعمال قید تصویر کردن (افکنش<sup>۳۷</sup>) ماتریس A روی PSD نیز متفاوت است. تصویر کردن مذکور در هر تکرار در الگوریتم ۱، هزینه محاسباتی زیادی دارد. نتایج آزمایش های انجام شده نشان می دهد که اگر بجای افکنش ماتریس A در هر تکرار، فقط ماتریس یادگیری نهایی روی PSD تصویر شود، علاوه بر کاهش هزینه محاسباتی، دقت آموزش نیز حفظ می شود [۳۱]. بنابراین در این مقاله نیز طبق الگوریتم ۲، از روش یکبار افکنش استفاده شده است. اندازه گام یادگیری در هر تکرار از فرآیند آموزش، در صورت افزایش مقدار تابع هزینه، با ضریب ۰/۵ کاهش و در صورت کاهش مقدار تابع هزینه با ضریب ۱/۲ افزایش یافته است.

#### ۴-۲ افزایش سرعت مقایسه با فاصله $\chi^2$ تقریبی

با کمی دقت در رابطه (۴)، می توان آن را به صورت رابطه (۲۲) نوشت:

الگوریتم ۲: حذف تصادفی به عنوان تنظیم کننده	
۱:	ورودی ها: پایگاه داده آموزشی $X \in \mathbb{R}^{D \times N}$ و گام یادگیری $\eta$
۲:	مقداردهی اولیه $A_0$ و $T$
۳:	حلقه تکرار $t = 1$ تا $T$
۴:	انتخاب نمونه تصادفی $\{\bar{x}_i, \bar{x}_j, \bar{x}_l\}_{j-i}$
۵:	ایجاد ماتریس تصادفی B
۶:	ایجاد پارامتر $\delta$ با استفاده از رابطه (۱۹)
۷:	$\hat{A}_{ij}^{t-1} = \delta_{ij} A_{ij}^{t-1}$ , $i, j = 1, 2, \dots, d$
۸:	$A_t = \hat{A}_{t-1} - \eta \nabla \ell$
۹:	اتمام حلقه تکرار
۱۰:	خروجی: $\prod_{psd}(\bar{A})$

$$A_t = (1-\beta)A_{t-1} - \eta \nabla \ell \quad (16)$$

در روش حذف تصادفی، اگر  $p(\delta_{ij} = 0) = \beta$  باشد، خواهیم داشت:

$$E[\hat{A}_{t-1}] = (1-\beta)A_{t-1} \quad (17)$$

با توجه به رابطه (۱۷)، به روزرسانی ماتریس A در الگوریتم SGD با روش حذف تصادفی بیان شده در رابطه (۱۳) معادل استفاده از تنظیم کننده نرم فروبنیوس در بهینه سازی (رابطه (۱۶)) است.

#### ۴-۲ تنظیم کننده ساختاریافته

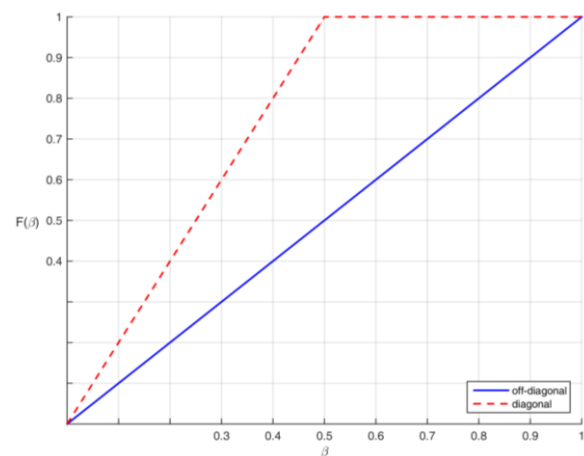
تنظیم کننده های متداول نظیر نرم فروبنیوس، به صورت ساده تمام عناصر ماتریس یادگیری را به یک شکل جریمه می کنند. اما واضح است که عناصر قطر اصلی ماتریس یادگیری از اهمیت بالاتری نسبت به سایر عناصر برخوردار هستند؛ چراکه عناصر قطر اصلی بیان کننده اهمیت هر یک از ویژگی ها بوده و سایر عناصر بیان کننده تعامل بین ویژگی ها هستند. در این بخش، روشی پیشنهاد می شود که تنظیم کننده متفاوتی را برای عناصر قطر اصلی ماتریس یادگیری و سایر عناصر در نظر می گیرد.

ماتریس تصادفی B که عناصر قطر اصلی این ماتریس از یک توزیع یکنواخت در بازه [0,0.5] و عناصر غیر قطر اصلی از یک توزیع یکنواخت در بازه [0,1] هستند را در نظر بگیرید. تابع توزیع تجمعی (CDF) برای عناصر این ماتریس در بازه [0,0.5] به صورت رابطه (۱۸) است؛ این توابع توزیع تجمعی در شکل ۱ نشان داده شده اند.

$$F(\beta) = \begin{cases} 2\beta & \text{for } B_{ij} : i = j \\ \beta & \text{for } B_{ij} : i \neq j \end{cases} \quad (18)$$

احتمال حذف تصادفی بر اساس ماتریس B به صورت رابطه (۱۹) تعریف می شود.

$$p(\delta_{ij} = 0) = p(B_{ij} \leq \beta) = \begin{cases} 2\beta & i = j \\ \beta & i \neq j \end{cases} \quad (19)$$



شکل ۱: تابع توزیع تجمعی برای عناصر قطر اصلی و غیر قطر اصلی ماتریس تصادفی B

طبق رابطه (۲۳) از  $\hat{x}_i$  و  $\hat{x}_j$  استفاده نمود. با این تقریب، رابطه (۸) به صورت رابطه (۲۵) قابل بازنویسی خواهد بود:

$$\chi_L^2(\hat{x}_i, \hat{x}_j) = \|L(\hat{x}_i - \hat{x}_j)\|^2 = \|L\hat{x}_i - L\hat{x}_j\|^2 \quad (25)$$

که در این رابطه،  $L$  همان ماتریس یادگیری است و  $\hat{x}_i$  در رابطه (۲۳) تعریف شده است. تمام  $L\hat{x}_i$  ها (که مربوط به داده‌های آموزشی هستند) به صورت غیربرخط و در مرحله استخراج ویژگی قابل محاسبه هستند. برای هر نمونه آزمایشی نیز فقط کافی است یکبار بردار  $L\hat{x}_j$  محاسبه شود و پس از آن می‌توان از معیار فاصله اقلیدسی برای تطبیق ویژگی‌ها استفاده کرده و سرعت محاسبات را نسبت به معیار فاصله  $\chi^2$  چند برابر نمود.

### ۳- پایگاه‌های داده و آزمایش‌ها

آزمایش‌هایی که در این بخش انجام شده است را می‌توان به دو دسته کلی تقسیم‌بندی نمود: (الف) ارزیابی الگوریتم یادگیری متریک پیشنهادی و (ب) ارزیابی دقت معیار  $\chi^2$  تقریبی پیشنهادی به منظور افزایش سرعت مقایسه داده‌ها. در این بخش به شرح پایگاه‌های داده و آزمایش‌های انجام شده بر روی آنها پرداخته می‌شود.

#### ۳-۴ پایگاه‌های داده

برای ارزیابی الگوریتم یادگیری متریک پیشنهادی از پایگاه‌های داده هیستوگرامی مختلف و پراستفاده در منابع یادگیری متریک استفاده شده است. برای ارزیابی رابطه  $\chi^2$  تقریبی پیشنهادی، علاوه بر پایگاه‌های داده هیستوگرامی، از سه پایگاه داده تصاویر چهره که برای شناسایی چهره و حالت چهره بکار می‌روند، استفاده شده است. در این بخش به معرفی این پایگاه‌های داده پرداخته می‌شود.

چهار پایگاه داده هیستوگرامی dslr، webcam، amazon و caltech که در [۳۲] معرفی شده‌اند، در اصل برای بازشناسی اشیاء مورد استفاده قرار می‌گیرند. این پایگاه‌های داده از ۱۰ شیء معمولی شامل کوله‌پشتی، دوچرخه، ماشین حساب، هدفون، صفحه‌کلید، رایانه لپ‌تاپ، نمایشگر، ماوس، لیوان و پروژکتور<sup>۲۸</sup> تشکیل شده است. داده‌های موجود در این پایگاه‌های داده، ویژگی‌های هیستوگرامی SURF<sup>۲۹</sup> به طول ۸۰۰ هستند. پایگاه‌های داده dslr، webcam، amazon و caltech به ترتیب دارای ۱۵۷، ۲۹۵، ۹۵۸ و ۱۱۲۳ نمونه هستند.

پایگاه داده corel دارای ۷۷۳ تصویر از ۱۰ دسته مختلف است. در هر دسته از این پایگاه داده، تعداد ۵۰ تا ۱۰۰ تصویر وجود دارد. برای این پایگاه داده نیز ویژگی‌های هیستوگرامی SIFT و CSIFT<sup>۳۰</sup> (به طول ۳۸۴) محاسبه و منتشر شده است<sup>۳۱</sup>، که در این مقاله از این ویژگی‌ها برای انجام آزمایش‌ها استفاده شده است. در پایگاه داده corel از مجموع ۷۷۳ داده، ۵۰ داده به عنوان مجموعه آزمایشی و مابقی داده‌ها به عنوان مجموعه آموزشی معرفی شده‌اند.

پایگاه داده optdigits از مجموعه KEEL<sup>۳۲</sup> [۳۴]، یکی دیگر از پایگاه‌های داده هیستوگرامی است که برای ارزیابی دقت دسته‌بندی

$$\chi^2(X, Y) = \sum_{b=1}^D \left( \frac{(\bar{x}_i)_b - (\bar{x}_j)_b}{\sqrt{(\bar{x}_i)_b + (\bar{x}_j)_b}} \right)^2 = \quad (22)$$

$$\sum_b \left( \frac{(\bar{x}_i)_b}{\sqrt{(\bar{x}_i)_b + (\bar{x}_j)_b}} - \frac{(\bar{x}_j)_b}{\sqrt{(\bar{x}_i)_b + (\bar{x}_j)_b}} \right)^2 = \sum_i (\hat{x}_i - \hat{x}_j)^2$$

که در آن،  $(\bar{x}_i)_b$  نشان دهنده عنصر  $b$  ام از هیستوگرام  $D$  بعدی  $\bar{x}_i$  است. رابطه (۲۲) برابر با مربع فاصله اقلیدسی بین  $\hat{x}_i$  و  $\hat{x}_j$  است و این دو عبارت خود از رابطه (۲۳) به دست می‌آیند:

$$\hat{x}_i = \bar{x}_i \oslash (\sqrt{\bar{x}_i \oplus \bar{x}_j}), \hat{x}_j = \bar{x}_j \oslash (\sqrt{\bar{x}_i \oplus \bar{x}_j}), \quad (23)$$

که در آنها  $\bar{x}_i$  و  $\bar{x}_j$  دو بردار (یا هیستوگرام) ورودی و نشان‌گرهای  $\oslash$ ،  $\oplus$  و  $\sqrt{\quad}$  به ترتیب نشان دهنده تقسیم، جمع و جذر درایه به درایه هستند.

حال اگر بتوان  $\hat{x}_i$  و  $\hat{x}_j$  را در مرحله استخراج ویژگی محاسبه نمود، طبق رابطه (۲۲)، در مرحله تطبیق ویژگی فقط کافی است که از معیار فاصله اقلیدسی استفاده نموده و سرعت تطبیق (مقایسه) ویژگی‌ها را افزایش داد؛ چراکه هزینه محاسباتی معیار فاصله اقلیدسی بسیار کمتر از معیار فاصله  $\chi^2$  است. اما به دلیل نیاز به وجود هر دو بردار ویژگی  $x_i$  و  $x_j$  در مخرج رابطه (۲۳)، نمی‌توان هر یک از عبارت‌های  $\hat{x}_i$  و  $\hat{x}_j$  را به صورت مستقل محاسبه نمود. با توجه به رابطه (۲۳)، می‌توان دریافت که مخرج  $\hat{x}_i$  و  $\hat{x}_j$  به نوعی یک نرمالیزاسیون عنصر به عنصر هیستوگرام‌ها است. این نرمالیزاسیون اثر تمام عناصر هیستوگرام را در فاصله نهایی تقریباً برابر نموده و از تحت تأثیر قرار گرفتن فاصله از عناصر بزرگ جلوگیری می‌کند. در این مقاله، این نرمالیزاسیون به صورت تقریبی و با استفاده از رابطه (۲۴) انجام شده است.

$$\hat{x}_i \cong \tilde{x}_i = \frac{x_i}{\sqrt{x_i + h_i}}, \hat{x}_j \cong \tilde{x}_j = \frac{x_j}{\sqrt{x_j + h_j}} \quad (24)$$

که در این رابطه،  $h$  یک بردار ویژگی هیستوگرامی مانند  $x$  است که به صورت تصادفی از بین داده‌های آموزشی انتخاب شده است. در بسیاری از مسائل شناسایی الگو که دارای یک پایگاه داده آموزشی (یا ثبت‌نام شده) بزرگ هستند، مدت زمان تطبیق ویژگی بسیار مهم‌تر از مدت زمان استخراج ویژگی است. چراکه در چنین مسائلی، استخراج ویژگی داده‌های آموزشی (ثبت‌نام) به صورت غیربرخط و فقط یکبار برای همیشه انجام می‌شود. بنابراین می‌توان بخش نرمالیزاسیون نشان داده شده در رابطه (۲۴) را به مرحله استخراج ویژگی منتقل نموده و سرعت تطبیق ویژگی را افزایش داد.

اگر بخواهیم از نتیجه یادگیری متریک پیشنهادی در الگوریتم ۲ برای مقایسه ویژگی‌ها استفاده کنیم، باید از رابطه (۸) استفاده نمود که محاسبه  $\bar{x}_{ioj}$  و  $\bar{x}_{joj}$  در آن هزینه محاسباتی بالایی خواهد داشت. از طرفی نمی‌توان این دو بردار را در مرحله استخراج ویژگی محاسبه نمود. برای کاهش این هزینه محاسباتی، به جای این دو بردار می‌توان

و  $\chi^2$  معیار فواصل  $\chi^2$  درجه دوم (QCS<sup>۲۶</sup> و QCN<sup>۳۳</sup>) [۳۳]، ITML [۹]، LMNN [۱۴]،  $\chi^2$ -LMNN [۲۰]، GB-LMNN<sup>۳۸</sup> [۲۱]، CDML<sup>۳۹</sup> [۳۸]، DML-eig<sup>۴۱</sup> [۱۵]، DMLMJ<sup>۴۱</sup> [۱۳]، NCML<sup>۴۲</sup> [۱۶] و PCML<sup>۴۳</sup> [۱۶] در جدول ۱ نشان داده شده است.

در آزمایش‌های مربوط به جدول ۱، برای پایگاه‌های داده dslr، amazon، webcam، caltech، همچون سایر منابع مقایسه شده، از روش اعتبارسنجی متقابل ۵ قسمتی<sup>۴۴</sup> با تقسیم‌بندی تصادفی استفاده شده است. برای پایگاه داده corel(SIFT) و corel (CSIFT) از تقسیم‌بندی ارائه شده در خود پایگاه داده (به ترتیب ۵۰ و ۷۲۳ داده به عنوان مجموعه آزمایشی و آموزشی) استفاده شده است. در پایگاه داده optdigits نیز از تقسیم‌بندی منتشر شده در پایگاه داده که به صورت اعتبارسنجی متقابل ۱۰ قسمتی است، استفاده شده است. بنابراین، به جز پایگاه داده corel که دارای یک مجموعه آزمایشی ثابت است، برای سایر پایگاه‌های داده، نتایج آزمایش‌ها براساس میانگین و انحراف معیار خطا گزارش شده است. تعداد همسایگی K در دسته‌بندی کننده KNN در تمامی این آزمایش‌ها برابر با ۳ قرار داده شده است. پیاده‌سازی QCS و QCN<sup>۴۵</sup>، ITML<sup>۴۶</sup>، LMNN<sup>۴۷</sup>، GB-LMNN<sup>۴۸</sup>، DML-eig<sup>۴۹</sup>، NCML، PCML<sup>۵۰</sup> منتشر شده و در دسترس است. در این روش‌ها، برای پایگاه‌های داده‌ای که مقاله‌ها گزارشی بر روی آن نداشته‌اند، از پیاده‌سازی‌های منتشر شده برای انجام آزمایش‌های جدید استفاده و نتایج آن در جدول ۱ گزارش شده است. اما برای روش‌های  $\chi^2$ -LMNN و CDML، به دلیل عدم انتشار پیاده‌سازی‌ها، آزمایش جدیدی انجام نگرفته و تنها نتایج گزارش شده توسط خود نویسندگان، در جدول ۱ ارائه شده است.

همان‌طور که در جدول ۱ ملاحظه می‌شود، خطای الگوریتم یادگیری متریک پیشنهادی در مقایسه با سایر روش‌ها کمتر است که این امر نشان دهنده کارایی این الگوریتم در دسته‌بندی داده‌های هیستوگرامی است.

در ادامه این بخش به ارزیابی رابطه  $\chi^2$  تقریبی پیشنهادی و مقایسه نتایج آن با رابطه  $\chi^2$  پرداخته می‌شود. برای این کار، ویژگی هیستوگرامی LBP<sup>۵۱</sup> [۳۹]، POEM<sup>۵۲</sup> [۴۰] و HOG [۴۱] از تصاویر پایگاه‌های داده CK+، JAFFE و Yale استخراج شده و شناسایی چهره و حالت چهره با استفاده از این ویژگی‌ها انجام شده است. نحوه استخراج این ویژگی‌ها به ترتیب زیر بوده است:

- LBP: هر تصویر به زیر نواحی  $6 \times 7$  تقسیم‌بندی شده است؛ سپس در هر زیرناحیه، هیستوگرام الگوهای یکنواخت LBP محاسبه شده است. در نتیجه طول بردار ویژگی نهایی برابر با  $2478 (59 \times 6 \times 7)$  بوده است.
- POEM: در یک تقسیم‌بندی مشابه با LBP و با در نظر گرفتن الگوهای یکنواخت، طول بردار ویژگی POEM چهار برابر LBP و برابر با ۹۹۱۲ بوده است.

کننده‌ها مورد استفاده قرار می‌گیرد. این پایگاه داده از ۵۶۲۰ تصویر از اعداد دست‌نویس تشکیل شده است. هر تصویر در پایگاه داده دارای یک بردار ویژگی به طول ۶۴ است؛ عناصر این بردار ویژگی، نشان‌دهنده تعداد پیکسل‌های روشن در نواحی مختلف تصاویر هستند. در optdigits، تقسیم‌بندی داده‌ها به مجموعه‌های آموزشی و آزمایشی به صورت اعتبارسنجی متقابل ۱۰ قسمتی<sup>۳۳</sup> منتشر شده است. پایگاه‌های داده هیستوگرامی dslr، amazon، webcam، caltech، corel و optdigits برای ارزیابی الگوریتم یادگیری متریک پیشنهادی استفاده شده است.

برای ارزیابی رابطه تقریبی پیشنهادی برای معیار فاصله  $\chi^2$  از پایگاه‌های داده CK+ [۳۵]، JAFFE [۳۶] و Yale [۳۷] استفاده شده است. پایگاه داده CK+ که برای شناسایی حالت چهره مورد استفاده قرار می‌گیرد، متشکل از ۱۲۳ شخص و ۷ حالت چهره است. پایگاه داده JAFFE از ۱۰ زن ژاپنی با ۷ حالت چهره مختلف تشکیل شده است و برای شناسایی چهره و حالت چهره مورد استفاده قرار می‌گیرد. پایگاه داده Yale نیز از ۱۵ شخص در شرایط نورپردازی مختلف تشکیل شده است و برای ارزیابی سیستم‌های شناسایی چهره استفاده می‌شود. تصاویر استفاده شده در این مقاله با استفاده از موقعیت دو چشم از تصاویر اصلی بریده شده و به اندازه  $110 \times 150$  پیکسل نرمالیزه شده‌اند. نمونه‌هایی از تصاویر این سه پایگاه داده چهره در شکل ۲ نشان داده شده است.

در آزمایش‌های مربوط به شناسایی چهره، یک تصویر از هر شخص به عنوان گالری<sup>۳۴</sup> و مابقی تصاویر به عنوان تصاویر پروب<sup>۳۵</sup> در نظر گرفته شده‌اند. در شناسایی حالت چهره نیز از روش اعتبارسنجی متقابل ۱۰ قسمتی به صورت مستقل از شخص برای تقسیم‌بندی داده‌های آموزشی و آزمایشی استفاده شده است.

### ۴-۳- آزمایش‌ها

در این قسمت، الگوریتم پیشنهادی مطرح شده در الگوریتم ۲، بر روی پایگاه‌های داده هیستوگرامی dslr، amazon، webcam، caltech، corel (SIFT)، corel (CSIFT) و optdigits آزمایش شده است. درصد خطای دسته‌بندی الگوریتم پیشنهادی در مقایسه با معیارهای فاصله اقلیدسی



شکل ۲: نمونه‌هایی از تصاویر پایگاه داده CK+ (ردیف بالا)، JAFFE (ردیف وسط) و Yale (ردیف پایین)



جدول ۱: مقایسه خطای دسته‌بندی روش‌های مختلف بر روی پایگاه‌های داده هیستوگرامی

روش	caltech	amazon	webcam	dslr	corel (CSIFT)	corel (SIFT)	optdigits
فاصله اقلیدسی	۵۳/۸ ± ۱/۳	۳۳/۷ ± ۰/۷	۴۳/۸ ± ۱/۷	۶۰/۶ ± ۳/۱	۵۰	۴۲	۱/۲ ± ۰/۴
فاصله $\chi^2$	۵۸/۸ ± ۱/۱	۳۴/۳ ± ۱/۰	۱۳/۰ ± ۱/۲	۲۲/۲ ± ۱/۸	۴۰	۴۰	۱/۳ ± ۰/۷
[۳۳] QCS	۵۷/۲ ± ۱/۲	۳۳/۹ ± ۲/۰	۱۹/۴ ± ۱/۱	۲۵/۶ ± ۲/۷	۳۸	۵۲	۱/۴ ± ۰/۶
[۳۳] QCN	۵۶/۱ ± ۱/۲	۳۴/۵ ± ۱/۵	۱۷/۵ ± ۲/۱	۲۷/۸ ± ۴/۱	۴۲	۵۲	۱/۳ ± ۰/۵
[۹] ITML	۵۲/۲ ± ۲/۱	۳۱/۶ ± ۱/۲	۱۲/۴ ± ۱/۶	۲۵/۰ ± ۳/۰	۴۲	۴۶	۱/۳ ± ۰/۶
[۱۴] LMNN	۵۰/۹ ± ۱/۴	۳۱/۸ ± ۱/۴	۱۵/۸ ± ۳/۰	۲۸/۹ ± ۱/۶	۳۶	۵۲	۱/۱ ± ۰/۴
[۲۰] $\chi^2$ -LMNN	۴۶/۵ ± ۱/۱	۲۳/۷ ± ۰/۸	۸/۳ ± ۰/۹	۲۰/۶ ± ۱/۱	-	-	-
[۲۰] GB-LMNN	۴۹/۸ ± ۱/۰	۲۹/۶ ± ۱/۷	۱۲/۴ ± ۰/۹	۲۲/۹ ± ۲/۷	۲۶	۵۸	۱/۱ ± ۰/۴
[۳۸] CDML	۴۲/۲ ± ۲/۴	۲۰/۸ ± ۳/۲	۵/۹ ± ۲/۳	۱۶/۷ ± ۴/۲	۲۶	۳۰	-
[۱۵] DML-eig	۵۵/۳ ± ۳/۷	۳۴/۳ ± ۳/۸	۲۰/۲ ± ۳/۴	۳۰/۱ ± ۴/۴	۴۶	۴۰	۱/۳ ± ۰/۵
[۱۳] DMLMJ	-	-	-	-	-	-	۱
[۱۶] NCML	۵۱/۶ ± ۱/۷	۳۲/۶ ± ۲/۱	۱۶/۷ ± ۲/۲	۲۸/۸ ± ۳/۹	۳۰	۳۴	۱/۳ ± ۰/۶
[۱۶] PCML	۵۲/۲ ± ۲/۵	۳۰/۲ ± ۳/۲	۱۵/۹ ± ۱/۱	۲۹/۱ ± ۳/۱	۲۸	۳۲	۱/۵ ± ۰/۴
روش پیشنهادی	۴۰/۶ ± ۴	۱۹/۸ ± ۱/۲	۵/۱ ± ۱/۲	۱۰/۸ ± ۲/۵	۲۴	۲۶	۰/۹ ± ۰/۴

ROC به دست آمده از این معیار و نیز معیار  $\chi^2$  در شناسایی حالت چهره در پایگاه داده CK+، در شکل ۳ نشان داده شده است. همان‌طور که در این شکل ملاحظه می‌شود، نمودارهای ROC دو معیار فاصله در ویژگی‌های مختلف بر همدیگر منطبق هستند که نشان دهنده دقت برابر این دو معیار است.

جدول ۲: شناسایی حالت چهره در پایگاه داده CK+ بر اساس POEM، LBP و HOG با استفاده از رابطه  $\chi^2$  و رابطه پیشنهادی

روش	HOG	POEM	LBP
فاصله $\chi^2$	۸۲/۲۶	۸۳/۷۹	۷۹/۲
فاصله پیشنهادی	۸۲/۱۴ ± ۰/۴۶	۸۳/۴۳ ± ۰/۳۲	۷۹/۲ ± ۰/۲۹

جدول ۳: شناسایی حالت چهره در پایگاه داده JAFFE بر اساس POEM، LBP و HOG با استفاده از رابطه  $\chi^2$  و رابطه پیشنهادی

روش	HOG	POEM	LBP
فاصله $\chi^2$	۹۱/۵۵	۹۲/۹۶	۹۰/۱۴
فاصله پیشنهادی	۹۱/۷۸ ± ۰/۲۵	۹۳/۰۵ ± ۰/۳	۹۰/۳۸ ± ۰/۵۱

جدول ۴: شناسایی چهره در پایگاه داده YALE بر اساس POEM، LBP و HOG با استفاده از رابطه  $\chi^2$  و رابطه پیشنهادی

روش	HOG	POEM	LBP
فاصله $\chi^2$	۷۳/۳۳	۷۶/۶۷	۷۷/۳۳
فاصله پیشنهادی	۷۲/۸ ± ۱/۰۳	۷۶/۹۳ ± ۱/۶۷	۷۷/۴۷ ± ۰/۸۲

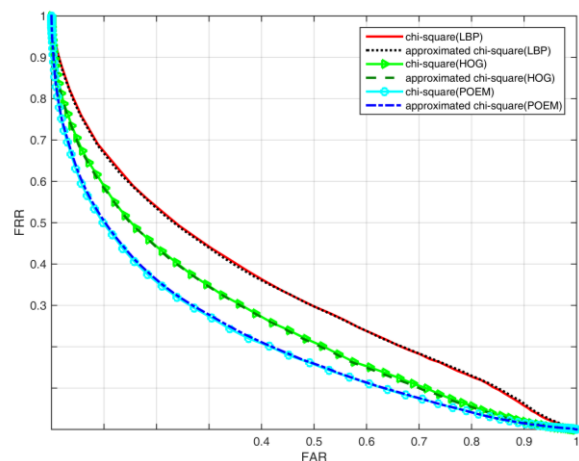
جدول ۵: شناسایی چهره در پایگاه داده JAFFE بر اساس POEM، LBP و HOG با استفاده از رابطه  $\chi^2$  و رابطه پیشنهادی

روش	HOG	POEM	LBP
فاصله $\chi^2$	۸۸/۱۸	۹۵/۵۷	۹۰/۶۴
فاصله پیشنهادی	۸۸/۶۷ ± ۰/۴	۹۵/۵۷ ± ۱/۰۱	۹۰/۸۹ ± ۰/۶۳

HOG برای محاسبه بردار ویژگی HOG، تصویر به زیر نواحی ۱۴ × ۱۹ تقسیم‌بندی شده است و در هر ناحیه، هیستوگرام مقادیر گرادیان در ۹ جهت محاسبه شده است. در نتیجه طول بردار ویژگی نهایی برابر با ۲۳۹۴ بوده است.

نتایج شناسایی حالت چهره در پایگاه‌های داده CK+ و JAFFE با استفاده از معیار فاصله  $\chi^2$  و نیز رابطه پیشنهادی به ترتیب در جدول ۲ و جدول ۳ نشان داده شده است. نتایج شناسایی چهره در پایگاه‌های داده Yale و JAFFE نیز به ترتیب در جدول ۴ و جدول ۵ آورده شده است. همان‌طور که از این نتایج مشاهده می‌شود، دقت رابطه پیشنهادی در شناسایی چهره و حالت چهره در پایگاه‌های داده مختلف تقریباً با رابطه  $\chi^2$  برابر است؛ که نشان دهنده کارایی مناسب رابطه پیشنهادی است.

یکی از روش‌های دقیق ارزیابی معیارهای مبتنی بر فاصله، محاسبه نمودار ROC<sup>۵۲</sup> است. برای ارزیابی دقیق‌تر رابطه پیشنهادی، نمودارهای



شکل ۳: نمودارهای ROC شناسایی حالت چهره در پایگاه داده CK+ با استفاده از رابطه فاصله  $\chi^2$  و نیز رابطه پیشنهادی

محاسباتی از مرتبه  $O(4N(D + D^2))$  خواهد و می توان گفت که هزینه محاسباتی روش پیشنهادی در مرحله دسته بندی داده ها نسبت به دیگر روش های یادگیری متریک که بر اساس فاصله مالهالانوبیس توسعه داده شده اند، در حدود ۲ برابر است. اما این هزینه محاسباتی بیشتر، با استفاده از رابطه تقریبی  $\chi^2$  پیشنهادی به صورت غیربرخط قابل انجام است. پیچیدگی محاسباتی دسته بندی با استفاده از رابطه تقریبی  $\chi^2$  پیشنهادی به همراه ماتریس یادگیری، از مرتبه  $O(2N(D + D^2) + 2D)$  است که به ازای  $N$  های بزرگ، تقریباً برابر با پیچیدگی محاسباتی روش های مبتنی بر فاصله مالهالانوبیس است.

#### ۴ جمع بندی

در این مقاله یک روش یادگیری متریک برای دسته بندی داده های هیستوگرامی با استفاده از دسته بندی کننده شناخته شده KNN پیشنهاد شد. برخلاف اکثر روش های یادگیری متریک که بر مبنای فاصله اقلیدسی (که برای فضای متعامد تعریف می شود) توسعه داده شده اند، در روش پیشنهادی، از فاصله  $\chi^2$  استفاده شده است. از نوآوری های این روش می توان به طرح یک تبدیل متغیر در رابطه فاصله  $\chi^2$  و تبدیل آن به فاصله اقلیدسی اشاره نمود که با به کارگیری آن می توان از روش های متداول یادگیری متریک مالهالانوبیس برای یادگیری متریک  $\chi^2$  بهره برد. این نکته باعث می شود که نسبت به KNN (بر اساس فاصله  $\chi^2$ ) و نیز سایر روش های یادگیری متریک موجود، بهبود قابل توجهی در دسته بندی داده های هیستوگرامی در پایگاه های داده مختلف حاصل شود. در الگوریتم یادگیری متریک پیشنهادی، به منظور جلوگیری از بیش تعلیم در یادگیری، یک تنظیم کننده حذف تصادفی جدید، طراحی و استفاده شده است. استفاده از این روش، که معمولاً در یادگیری عمیق به کار می رود، باعث می شود که با یک پیاده سازی نسبتاً ساده بتوان قدرت تعمیم الگوریتم را افزایش داد. از نوآوری های دیگری این مقاله می توان به معرفی یک رابطه تقریبی پیشنهادی برای فاصله  $\chi^2$  اشاره کرد که منجر به افزایش سرعت محاسبات در مرحله مقایسه بردارهای ویژگی می شود. با استفاده از این تقریب، بخش بزرگی از محاسبات مربوط به فاصله  $\chi^2$  در مرحله استخراج ویژگی و به صورت غیربرخط قابل محاسبه بوده و در مرحله مقایسه ویژگی ها به راحتی می توان از معیار فاصله اقلیدسی استفاده نمود. بر اساس آزمایش های متنوع انجام شده، این رابطه تقریبی با حفظ دقت معیار فاصله  $\chi^2$ ، سرعت آن را در حدود ۲/۵ برابر افزایش می دهد. همچنین این رابطه بدون کاهش دقت مقایسه، در کنار الگوریتم یادگیری متریک پیشنهادی نیز قابل استفاده بوده است.

#### پوست الف: محاسبه گرادیان

برای محاسبه گرادیان تابع هزینه داده شده در رابطه (۱۰)، فرض کنید که  $H_{ij}$  به صورت  $H_{ij} = (\vec{x}_{ioj} - \vec{x}_{joi})(\vec{x}_{ioj} - \vec{x}_{joi})^T$  تعریف شود که بردارهای  $\vec{x}_{ioj}$  و  $\vec{x}_{joi}$  پیش تر برای رابطه (۷) تعریف شدند؛ با این

به منظور ارزیابی دقت معیار تقریبی  $\chi^2$  پیشنهادی در کنار الگوریتم یادگیری متریک شرح داده شده، داده های هیستوگرامی پایگاه های داده amazon، webrtc، dslr و caltech با استفاده از رابطه (۸) و رابطه (۲۵) (که به ترتیب رابطه اصلی  $\chi^2$  و رابطه تقریبی  $\chi^2$  پیشنهادی را با ماتریس یادگیری تلفیق کرده اند) مشابه با آزمایش های مربوط به جدول ۱ دسته بندی شده اند. نتایج حاصل از این آزمایش ها در جدول ۶ مقایسه شده است. همان طور که در این جدول مشاهده می شود، با استفاده از رابطه  $\chi^2$  تقریبی، دقت الگوریتم یادگیری متریک پیشنهادی حفظ شده است.

برای ارزیابی هزینه محاسباتی رابطه  $\chi^2$  تقریبی پیشنهادی و مقایسه آن با  $\chi^2$  از پیاده سازی این دو رابطه در نرم افزار MATLAB و در یک رایانه شخصی معمولی core i5 با فرکانس کلک ۲/۶۷ GHz استفاده شده است. جدول ۷، تعداد تطبیق های انجام شده در واحد ۱ ثانیه را برای رابطه  $\chi^2$  (رابطه (۴)) و رابطه  $\chi^2$  تقریبی بر اساس روابط (۲۲) و (۲۴) با استفاده از بردارهای ویژگی مختلف نشان می دهد. همان طور که از مقایسه نتایج در این جدول ملاحظه می شود، سرعت تطبیق با استفاده از رابطه  $\chi^2$  تقریبی پیشنهادی در حدود ۲/۵ برابر رابطه  $\chi^2$  است که نشان دهنده برتری قابل توجه آن از نظر هزینه محاسباتی است.

در صورتی که ابعاد و تعداد داده های آموزشی به ترتیب برابر با  $D$  و  $N$  باشد، پیچیدگی محاسباتی دسته بندی یک داده آزمایشی با استفاده از فاصله اقلیدسی از مرتبه  $O(2ND)$  خواهد بود. حال اگر ماتریس یادگیری  $A$  مربعی و با ابعاد  $D \times D$  باشد و از فاصله مالهالانوبیس برای دسته بندی داده ها استفاده شود، این پیچیدگی محاسباتی از مرتبه  $O(2N(D + D^2))$  خواهد بود؛ چرا که تعداد عناصر ماتریس  $A$  برابر با  $D^2$  بوده و بر اساس رابطه (۱)، باید یک ضرب ماتریسی در محاسبات لحاظ شود. برای فاصله  $\chi^2$ ، به دلیل وجود یک جمع (در مخرج) و یک تقسیم اضافه نسبت به فاصله اقلیدسی، این پیچیدگی دو برابر فاصله اقلیدسی و از مرتبه  $O(4ND)$  است. بنابراین، در دسته بندی با استفاده از رابطه پیشنهادی (رابطه (۷)) با در نظر گرفتن ماتریس  $A$ ، پیچیدگی

جدول ۶: مقایسه دقت دسته بندی داده های هیستوگرامی با روش یادگیری متریک پیشنهادی با استفاده از رابطه  $\chi^2$  و رابطه  $\chi^2$  تقریبی

پایگاه داده	$\chi^2$ (رابطه (۸))	$\chi^2$ تقریبی (رابطه (۲۵))
dslr	$10/8 \pm 2/5$	$10/9 \pm 2/6$
webrtc	$5/1 \pm 1/2$	$5/1 \pm 1/1$
amazon	$19/8 \pm 1/2$	$19/7 \pm 1/2$
caltech	$40/6 \pm 4$	$40/7 \pm 3/5$

جدول ۷: تعداد مقایسه های انجام شده در واحد ثانیه با استفاده از رابطه  $\chi^2$  و رابطه پیشنهادی

روش	HOG	POEM	LBP
فاصله $\chi^2$	۱۵۶۲۵	۳۳۱۰	۱۲۸۲۰
فاصله $\chi^2$ تقریبی	۳۷۰۳۰	۹۱۷۵	۳۲۲۲۵

information,” In Advances in Neural Information Processing Systems (NIPS), vol. 15, pp. 521-528, 2003.

- [8] M. Schultz and T. Joachims, “Learning a distance metric from relative comparisons,” in Advances in Neural Information Processing Systems (NIPS), vol. 16, pp. 41-48, 2004.
- [9] J. V. Davis, B. Kulis, P. Jain, S. Sra and I. S. Dhillon, “Information theoretic metric learning,” In Proceedings of the 24<sup>th</sup> International Conference on Machine Learning (ICML), pp. 209-216, 2007.
- [10] G. Kunapuli and J. Shavlik, “Mirror descent for metric learning: a unified approach,” In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database (ECML/PKDD), pp. 859-874, 2012.
- [11] Q. Qian, R. Jin, J. Yi, L. Zhang and S. Zhu, “Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (SGD),” Machine Learning, vol. 99, no. 3, pp. 353-372, 2015.
- [12] M. T. Law, N. Thome and M. Cord, “Learning a distance metric from relative comparisons between quadruplets of images,” International Journal of Computer Vision, vol. 121, no. 1, pp. 65-94, 2017.
- [13] B. Nguyen, C. Morell and B. D. Baets, “Supervised distance metric learning through maximization of the Jeffrey divergence,” Pattern Recognition, vol. 64, pp. 215-225, 2017.
- [14] K. Q. Weinberger and K. S. Lawrence, “Distance metric learning for large margin nearest neighbor classification,” Journal of Machine Learning Research, vol. 10, pp. 207-244, 2009.
- [15] Y. Ying and P. Li, “Distance metric learning with eigenvalue optimization,” Journal of Machine Learning Research, vol. 13, no. 1, pp. 1-26, Jan. 2012.
- [16] W. Zuo, F. Wang, D. Zhang, L. Lin, Y. Huang, D. Meng and L. Zhang, “Distance metric learning via iterated support vector machines,” IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4937-4950, 2017.
- [17] Z. Hao, Y. Ruan, Y. Xiao and B. Liu, “A multi-task-based classification framework for multi-instance distance metric learning,” Neurocomputing, vol. 275, pp. 418-429, 2018.

[۱۸] وحیده منعمی‌زاده و جواد حمیدزاده، «جستجوی k نزدیک‌ترین همسایه تقریبی با روش ترکیب خطی»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۷، شماره ۳، صفحات ۱۲۴۹-۱۲۳۷، پاییز ۱۳۹۶.

[۱۹] مهرداد حیدری ارجلو، سید قدرت‌اله سیفالسادات و مرتضی رزاز، «یک روش هوشمند تشخیص جزیره در شبکه‌ی توزیع دارای تولیدات پراکنده مبتنی بر تبدیل موجک و نزدیک‌ترین K-همسایگی (KNN)»، مجله مهندسی برق دانشگاه تبریز، جلد ۱۹، شماره ۱، صفحات ۱۵-۲۶، ۱۳۹۲.

[20] D. Kedem, S. Tyree, F. Sha, G. R. Lanckriet and K. Q. Weinberger, “Non-linear metric learning,” In Advances in Neural Information Processing Systems (NIPS), vol. 25, pp. 2573-2581, 2012.

[۲۱] نصیبه اسدی‌پور ماسوله و اسدالله شاه‌بهرامی، «تخمین خودکار سن از روی تصویر چهره با تلفیق ویژگی‌های آماری و بافت»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۷، شماره ۳، صفحات ۸۴۲-۸۲۹، پاییز ۱۳۹۶.

- [22] T. Ahonen, A. Hadid and M. Pietikäinen, “Face recognition with local binary patterns,” In European Conference on Computer Vision (ECCV-2004), pp. 469-481, Springer, Berlin, Heidelberg, 2004.
- [23] G. Zhang, X. Huang, S. Z. Li, Y. Wang and X. Wu, “Boosting local binary pattern (LBP)-based face recognition,” In Advances in Biometric Person Authentication, pp. 179-186, Springer Berlin Heidelberg, 2004.
- [24] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” Arxiv Preprint Arxiv, pp. 1207-0580, 2012.

تعریف می‌توان رابطه (Y) را به صورت  $\chi_A^2(\vec{x}_i, \vec{x}_j) = \text{tr}(A H_{ij})$  بازنویسی کرد. در نتیجه می‌توان تابع هزینه داده شده در رابطه (۱۰) را به صورت رابطه (۲۶) نوشت.

$$\ell(A, X) = (1 - \mu) \sum_{i,j-i} \text{tr}(A H_{ij}) + \mu \sum_{i,j-i, l: y_i \neq y_l} [1 + \text{tr}(A H_{ij}) - \text{tr}(A H_{il})]_+ \quad (26)$$

فرض کنید که  $N$  مجموعه‌ای از داده‌های آموزشی سه‌تایی<sup>۵۴</sup>  $(i, j, l)$  باشد که تابع اتلاف لولایی موجود در عبارت سمت راست رابطه (۲۶) را تحریک می‌کنند؛ در این صورت گرادیان رابطه (۲۶) به صورت رابطه (۲۷) محاسبه می‌شود.

$$G = \frac{\partial \ell}{\partial A} = (1 - \mu) \sum_{i,j-i} H_{ij} + \mu \sum_{(i,j,l) \in N} (H_{ij} - H_{il}) \quad (27)$$

در فرآیند یادگیری و محاسبه گرادیان در هر تکرار از الگوریتم، می‌توان از محاسبات انجام شده در تکرار قبل استفاده نموده و هزینه محاسبات را به‌طور قابل ملاحظه‌ای کاهش داد [۱۴]. برای این کار، گرادیان تکرار مرحله  $t + 1$  ام  $(G_{t+1})$  را می‌توان طبق رابطه (۲۸) بر اساس گرادیان تکرار مرحله  $t$  ام  $(G_t)$  و تغییرات بین مجموعه  $N_t$  و  $N_{t+1}$  محاسبه نمود:

$$G_{t+1} = G_t - \mu \sum_{(i,j,l) \in N_t - N_{t+1}} (H_{ij} - H_{il}) + \mu \sum_{(i,j,l) \in N_{t+1} - N_t} (H_{ij} - H_{il}) \quad (28)$$

که  $H$  پیش‌تر برای بیان رابطه (۲۶) تعریف شد و عبارت  $(i, j, l) \in N_t - N_{t+1}$  بیان کننده داده‌های سه‌تایی است که عضو مجموعه  $N_t$  هستند اما در  $N_{t+1}$  وجود ندارند. در این رابطه، اثر داده‌های سه‌تایی که دیگر فعال نیستند از گرادیان کاسته شده و اثر داده‌های سه‌تایی که فعال شده‌اند به گرادیان افزوده می‌شود. این روش، به‌خصوص در زمانی که گام یادگیری کوچک بوده و تغییرات  $N_{t+1}$  نسبت به  $N_t$  کم است، هزینه محاسباتی را بسیار کاهش می‌دهد.

## مراجع

- [1] B. Kulis, “Metric learning: a survey,” Foundations and Trends® in Machine Learning, vol. 5, no. 4, pp. 287-364, 2013.
- [2] A. Bellet, A. Habrard and M. Sebban, “A survey on metric learning for feature vectors and structured data,” ArXiv Preprint ArXiv, pp. 1306.6709, 2013.
- [3] C. Jin and S. W. Jin, “Image distance metric learning based on neighborhood sets for automatic image annotation,” Journal of Visual Communication and Image Representation, vol. 34, pp. 167-175, 2016.
- [4] J. Hu, J. Lu and Y. Tan, “Sharable and individual multi-view metric learning,” IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 99, 2017 (In Press).
- [5] H. Yan, “Kinship verification using neighborhood repulsed correlation metric learning,” Image and Vision Computing, vol. 60, no. 1, pp. 91-97, 2017.
- [6] H. Yan, and J. Hu, “Video-based kinship verification using distance metric learning,” Pattern Recognition, vol. 75, pp.15-24, 2018.
- [7] E. P. Xing, M. I. Jordan, S. J. Russell and A. Y. Ng, “Distance metric learning with application to clustering with side-

- and F. Herrera, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307-318, 2009.
- [35] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression," *IEEE Computer Society Conf. Computer Vision and Pattern Recognition-Workshops*, pp. 94-101, June 2010,
- [36] M. J. Lyons, J. Budynek and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, 1999.
- [37] P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [38] W. Yang, L. Xu, X. Chen, F. Zheng and Y. Liu, "Chi-squared distance metric learning for histogram data," *Mathematical Problems in Engineering*, vol. 2015, 2015.
- [39] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.
- [40] N. S. Vu and A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1352-1365, 2012.
- [41] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In *2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886-893, June 2005.
- [25] L. V. D. Maaten, M. Chen, S. Tyree and K. Q. Weinberger, "Learning with marginalized corrupted features," In *International Conference on Machine Learning (ICML)*, pp. 410-418, 2013.
- [26] S. Wager, S. Wang and P. Liang, "Dropout training as adaptive regularization," In *Advances in Neural Information Processing Systems (NIPS)*, vol. 26, pp. 351-359, 2013.
- [27] Q. Qian, J. Hu, R. Jin, J. Pei and S. Zhu, "Distance metric learning using dropout: a structured regularization approach," In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 323-332, 2014.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [29] B. Shaw, B. C. Huang and T. Jebara, "Learning a distance metric from a network," In *Advances in Neural Information Processing Systems (NIPS)*, vol. 24, pp. 1899-1907, 2011.
- [30] Q. Qian, R. Jin, S. Zhu and Y. Lin, "Fine-grained visual categorization via multi-stage metric learning," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3716-3724, 2015.
- [31] G. Chechik, V. Sharma, U. Shalit and S. Bengio, "Large scale online learning of image similarity through ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109-1135, 2010.
- [32] K. Saenko, B. Kulis, M. Fritz and T. Darrell, "Adapting visual category models to new domains," In *European Conference on Computer Vision (ECCV-2010)*, pp. 213-226, 2010.
- [33] O. Pele and M. Werman, "The quadratic-chi histogram distance family," In *European Conference on Computer Vision (ECCV-2010)*, pp. 749-762. Springer, Berlin, Heidelberg, 2010.
- [34] J. Alcalá-Fdez, L. Sánchez, S. García, M. D. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández

<sup>29</sup> Speeded-Up Robust Features (SURF)

<sup>30</sup> Color-SIFT

<sup>31</sup> <http://ofirpele.droppages.com/>

<sup>32</sup> Knowledge Extraction based on Evolutionary Learning (KEEL); <http://sci2s.ugr.es/keel/datasets.php>

<sup>33</sup> 10-fold cross validation

<sup>34</sup> gallery

<sup>35</sup> probe

<sup>36</sup> Quadratic-Chi-Squared (QCS)

<sup>37</sup> Quadratic-Chi-Normalized (QCN)

<sup>38</sup> Gradient Boosted-LMNN (GB-LMNN)

<sup>39</sup> Chi-Squared Distance Metric Learning (CDML)

<sup>40</sup> Distance Metric Learning with Eigenvalue Optimization

<sup>41</sup> Distance Metric Learning through Maximization of the Jeffrey divergence

<sup>42</sup> Nonnegative-coefficient Constrained Metric Learning

<sup>43</sup> Positive-semidefinite Constrained Metric Learning

<sup>44</sup> 5-fold cross validation

<sup>45</sup> <http://www.ariel.ac.il/sites/ofirpele/QC/code/>

<sup>46</sup> <http://www.cs.utexas.edu/users/pjain/itml/>

<sup>47</sup> <http://www.cse.wustl.edu/~kilian/code/files/mLMMN2.4.zip>

<sup>48</sup> <https://www.cs.cornell.edu/~kilian/code/code.html>

<sup>49</sup> <http://empslocal.ex.ac.uk/people/staff/vy267/dml-eig-copy.zip>

<sup>50</sup> <https://github.com/csfwang/ISVM>

<sup>51</sup> Local Binary Patterns (LBP)

<sup>52</sup> Patterns of Oriented Edge Magnitudes (POEM)

<sup>53</sup> Receiver Operating Characteristic (ROC)

<sup>54</sup> triplet

<sup>1</sup> offline

<sup>2</sup> class

<sup>3</sup> clustering

<sup>4</sup> classification

<sup>5</sup> Automatic image annotation

<sup>6</sup> Mahalanobis

<sup>7</sup> Positive Semi-Definite (PSD)

<sup>8</sup> Information-Theoretic Metric Learning (ITML)

<sup>9</sup> divergence

<sup>10</sup> kernel classification

<sup>11</sup> Support Vector Machine (SVM)

<sup>12</sup> Large Margin Nearest Neighbors (LMNN)

<sup>13</sup> K-Nearest Neighbors (KNN)

<sup>14</sup> Chi-square

<sup>15</sup> regularizer

<sup>16</sup> over-fitting

<sup>17</sup> simplex

<sup>18</sup> Histogram of Oriented Gradients (HOG)

<sup>19</sup> Stochastic Gradient Descent (SGD)

<sup>20</sup> dropout

<sup>21</sup> targets neighbors

<sup>22</sup> impostors

<sup>23</sup> hinge loss

<sup>24</sup> semidefinite programming (SDP)

<sup>25</sup> slack variables

<sup>26</sup> Frobenius norm

<sup>27</sup> projection

<sup>28</sup> back pack, bike, calculator, headphones, keyboard, laptop computer, monitor, mouse, mug, and projector