ارائه یک روش مؤثر برای یادگیری مقاوم متریک در برابر نوفهٔ برچسب

داود ذبیح‌زاده خواجوی[1*]، سعید زاهدی[2] و رضا منصفی[3]

[1] گروه مهندسی کامپیوتر، دانشگاه حکیم سبزواری، سبزوار، ایران

[2و3] دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

چکیده

تعیین شباهت/ فاصله داده‌ها در بسیاری از الگوریتم‌های یادگیری ماشین، شناسایی الگو و داده‌کاوی کاربرد دارد. در بسیاری از کاربردها، معیارهای عمومی شباهت/فاصله کارایی بالایی ندارد و به‌طورمعمول با استفاده از داده‌ها می‌توان معیار مناسب‌تری را یاد گرفت. داده‌های آموزشی برای این منظور به‌طورمعمول به‌صورت زوج‌های مشابه و نامشابه و یا محدودیت‌های سه‌گانه هستند. در کاربردهای واقعی، این داده‌های آموزشی از طریق اینترنت و به‌طورمعمول با روش‌هایی نظیر Crowdsourcing جمع‌آوری می‌شود که می‌تواند حاوی نوفه و اطلاعات اشتباه باشد. کارایی روش‌های یادگیری متریک در صورت وجود اطلاعات آموزشی نوفه‌ای و اشتباه به شدت افت می‌کند و حتی ممکن است این روش‌ها از معیارهای عمومی فاصله نظیر اقلیدسی نیز بدتر عمل کنند. بنابراین نیاز به مقاوم‌سازی روش‌های یادگیری متریک در برابر نوفه برچسب وجود دارد. در این پژوهش، یک تابع احتمالاتی جدید برای تعیین احتمال نوفه‌ای‌بودن برچسب داده‌ها با استفاده از محدودیت‌های سه‌گانه آموزشی ارائه‌شده است که باعث می‌شود، الگوریتم یادگیری متریک بتواند داده‌های پرت و نوفه‌ای را شناسایی کند و تأثیر آن‌ها را فرایند یادگیری کاهش دهد. همچنین نشان داده شده است که چگونه از اطلاعات به‌دست‌آمده می‌توان برای افزایش کارایی الگوریتم مبتنی بر متریک (مانند kNN) بهره برد و عملکرد آن را به‌طور قابل‌ملاحظه‌ای افزایش داد. نتایج آزمایش‌ها بر روی مجموعه‌ای از داده‌های ساختگی و واقعی، تأیید می‌کند که روش پیشنهادی به‌طور قابل‌ملاحظه‌ای کارایی روش‌های یادگیری متریک را در محیط‌هایی با نوفه برچسب بهبود می‌بخشد و بر روش‌های همتا در مرزهای دانش در سطوح مختلف نوفه برچسب برتری دارد.

واژگان کلیدی: یادگیری متریک مقاوم، نوفه برچسب، داده‌های پرت، معیار فاصله

An Effective Approach for Robust Metric Learning in the Presence of Label Noise

Davood Zabihzadeh[*1], Saeed Zahedi[2] & Reza Monsefi[3]

[1]Department of Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran

[2,3]Engineering Faculty, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

Many algorithms in machine learning, pattern recognition, and data mining are based on a similarity/distance measure. For instance, the kNN classifier and clustering algorithms such as k-means require a similarity/distance function. Also, in Content-Based Information Retrieval (CBIR) systems, we need to rank the retrieved objects based on the similarity to the query. As generic measures like Euclidean and cosine similarity are not appropriate in many applications, metric learning algorithms have been developed with the aim of learning an optimal distance function from data. These methods often need training data in the form of pair or triplet sets. Nowadays, this training data is popularly obtained via crowdsourcing from the Internet. Therefore, this information may be contaminated with label noise resulting in the poor performance of the learned metric. In some datasets, even it is possible that the learned metrics perform worse than the general ones such as Euclidean. To address this emerging challenge, we present a new robust metric learning algorithm that can identify outliers and label noise simultaneously from training side information. For this purpose, we model the probability distribution of label noise based on information in the training data. The proposed distribution function

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۳۹۸/۹/۱۴ • تاریخ پذیرش: ۱۳۹۹/۵/۲۸ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]

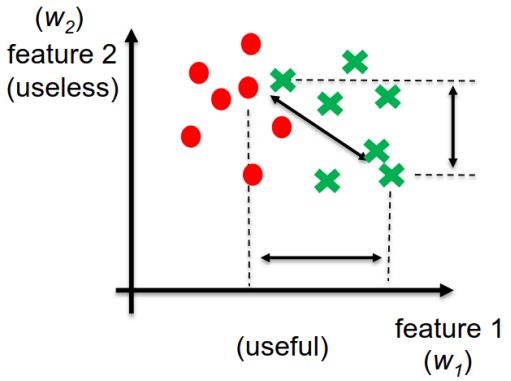
efficiently assigns the high probability to the data points contaminated with label noise. On the other hand, its value on the normal instances is near zero. Afterward, we weight the training instances according to these probabilities in our metric learning optimization problem. The proposed optimization problem can be solved using available SVM libraries such as LibSVM efficiently. Note that the proposed approach for identifying data with label noise is general and can easily be applied to any existing metric learning algorithms. After the metric learning phase, we utilized both the weights and the learned metric to enhance the accuracy of the metric-based classifier such as kNN. Several experiments are conducted on both real and synthetic datasets. The results confirm that the proposed algorithm enhances the performance of the learned metric in the presence of label noise and considerably outperforms state-of-the-art peer methods at different noise levels.

Keywords: Robust Metric Learning, Label Noise, Outlier, Distance Measure

۱- مقدمه

معیار شباهت/ فاصله در بسیاری از مسائل یادگیری ماشین و شناسایی الگو کاربرد دارد. برای مثال الگوریتم‌های خوشه‌بندی مانند k-means و یا طبقه‌بند kNN نیاز به معیاری برای سنجش فاصله بین داده‌ها دارند؛ همچنین در بازیابی اطلاعات برحسب محتوا (Content Based Information Retrieval)، باید اشیا بازیابی شده برحسب میزان شباهت به شیء پرس‌وجو مرتب شوند. در بسیاری از کاربردها، معیارهای عمومی شباهت/فاصله مانند فاصله اقلیدسی و یا شباهت کسینوسی عملکرد مناسبی ندارند. برای مثال شکل (۱) ضعف معیار فاصله اقلیدسی را نشان می‌دهد. همان‌طور که مشاهده می‌شود در شکل (۱) ویژگی نخست برای تمایز داده‌های دو طبقه مفید و ویژگی دوم نامفید است؛ درحالی‌که اهمیت هر دو ویژگی در معیار فاصله اقلیدسی یکسان است.

در بیشتر کاربردها می‌توان از روی‌داده‌ها، معیار شباهت/فاصله مؤثرتری را یادگیری گرفت. الگوریتم‌های یادگیری متریک از محدودیت‌های زوج و یا سه‌گانه‌های آموزشی استفاده می‌کنند که به‌صورت زیر تعریف می‌شود:



(شکل-۱): ضعف معیار فاصله اقلیدسی، این معیار برای سنجش فاصله داده‌ها به همه ویژگی‌ها وزن یکسانی می‌دهد؛ درحالی‌که ویژگی w_1 در این شکل برای تمایز داده‌های دو کلاس اهمیت بیشتری دارد.

(Figure-1): Weakness of Euclidean distance measure. Here, the w_1 is a more discriminative feature compared to w_2. However, Euclidean assigns the same weight to each feature.

$$S = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) \mid \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ should be similar}\}$$
$$D = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) \mid \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ should be dissimilar}\}$$
$$T = \{(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) \mid \boldsymbol{x}_i \text{ should be more closer to } \boldsymbol{x}_j \text{ than to } \boldsymbol{x}_k\}$$

در روابط بالا، S و D به‌ترتیب مجموعه زوج‌های مشابه و نامشابه نام دارد و T مجموعه سه‌گانه‌ها (Triplets) را نشان می‌دهد که به آن محدودیت‌های نسبی هم گفته می‌شود.

تاکنون بیشترین توجه در یادگیری متریک به فاصله ماهالونوبیس اختصاص داده شده است. البته معیارهای جدیدتر مانند تعریف شباهت بر اساس توزیع داده‌ها [1] و یا شکل کلی‌تر از فاصله ماهالونوبیس [2] نیز ارائه شده است. فاصله ماهالونوبیس به‌صورت زیر تعریف می‌شود:

$$d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_j)^2 = (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \boldsymbol{M}(\boldsymbol{x}_i - \boldsymbol{x}_j) \qquad (۱)$$

در رابطه بالا $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ (d بعد داده‌ها را نشان می‌دهد) یک ماتریس نیمه معین مثبت (p.s.d[1]) است که باید از محدودیت‌های آموزشی یاد گرفته شود.

به‌طورکلی در هر الگوریتمی که از معیار فاصله استفاده می‌کند، استفاده از یادگیری متریک اهمیت ویژه‌ای می‌یابد. برای نمونه دقت طبقه بند مبتنی بر فاصله kNN را می‌توان با استفاده از روش‌های یادگیری متریک بهبود بخشید. در بینایی ماشین از یادگیری متریک برای پیداکردن یک فضای مقاوم در برابر ویژگی‌هایی مانند تغییر موقعیت، شرایط نوری و زاویه‌ها استفاده می‌شود. برای مثال، از یادگیری متریک در کارهای شناسایی چهره [3]، شناسایی مجدد شخص[2] [2 , 4]، ردیابی بصری [5]، حاشیه‌نویسی/شرح نگاری تصویر [6] استفاده شده است. از کاربردهای دیگر یادگیری متریک می‌توان به سامانه‌های بازیابی اطلاعات برحسب محتوا (CBIR[3]) اشاره کرد. در CBIR هدف بازیابی فهرستی مرتب از اشیا است که بیش‌ترین شباهت را به شیء مورد جستجو[4] دارند [7-10]. برای مشاهده موارد بیشتر به [11] مراجعه کنید.

[1] Positive Semi Definite
[2] person re-identification
[3] Content Based Information Retrieval
[4] Query

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

هرچند یادگیری متریک در کاربردهای مختلف موفق بوده ولی یکی از چالش‌های مهم آن، داده‌های پرت و یا وجود نوفه برچسب است که باعث افت کارایی قابل‌ملاحظه آن می‌شود [12]. برای حل این مشکل، الگوریتم‌های یادگیری متریک مقاوم ارائه شده‌اند. بیشتر روش‌های فعلی در این زمینه، روش‌های احتمالاتی هستند که احتمال نوفه‌ای‌بودن محدودیت‌های آموزشی را در حین اجرای الگوریتم تقریب می‌زنند. در ادامه این روش‌ها بررسی می‌شوند.

GMENs[1] [13] روش یادگیری متریک باوجود زوج‌های آموزشی نوفه‌ای است. در این مقاله ابتدا یادگیری متریک در حالت عادی به‌صورت یک مسأله دو طبقه فرموله شده و سپس روش ارائه‌شده برای داده‌های نوفه‌ای تعمیم یافته است.

در RML[2] [14] یک چارچوب عمومی برای تشخیص محدودیت‌های نوفه‌ای در حین فرایند یادگیری متریک ارائه شده است. این الگوریتم فرض می‌کند درصد اطلاعات نوفه‌ای از پیش مشخص است، ولی در مورد اینکه کدام اطلاعات اشتباه هستند، اطلاعی نداریم. در این روش، ابتدا یادگیری متریک به‌عنوان مسأله بهینه‌سازی عدد صحیح ترکیبی فرموله شده است که پیچیدگی زمانی نمایی دارد؛ سپس این مسأله با یک مسأله محدب-مقعر تقریب زده شده و برای حل آن از بهینه‌سازی هموار Nesterov [15] استفاده شده است.

RNCA[3] [16] برای مقاوم‌سازی روش NCA (Goldberger, Hinton et al. 2005) در برابر نوفه برچسب ارائه‌شده است. برای این منظور، تأثیر برچسب نوفه‌ای بر روی مشتق تابع درست‌نمایی (likelihood) بررسی‌شده و این الگوریتم با مدل‌سازی احتمال درست‌بودن برچسب، سعی می‌کند تأثیر آن را کاهش یابد.

در [17] برای یادگیری متریک مقاوم در برابر نوفه، تابع هدف مشابه [18] با استفاده از فاصله نُرم ۱ پیشنهاد شده است. بااین‌وجود، مسأله ارائه‌شده به مقداردهی اولیه متریک بسیار حساس است، ازاین‌رو در آزمایش‌ها، از جواب حاصل از الگوریتم [18] برای مقداردهی اولیه متریک استفاده شده است. همچنین در تابع نُرم ۱، ضرر[4] به‌صورت خطی رُشد می‌کند؛ بنابراین، این تابع ضرر چندان برای مجموعه‌داده‌های حاوی نوفهٔ برچسب و داده‌های پرت مناسب نیست.

[1] Generalized Maximum Entropy Model for learning from Noisy side information
[2] Robust Metric Learning
[3] Robust *Robust* Neighbourhood Component Analysis
[4] Loss

از کارهای جدید درزمینهٔ یادگیری متریک مقاوم، استفاده از استنتاج بیزی [19,20] است. در این مقاله‌ها بجای تخمین نقطه‌ای ماتریس فاصله که باعث حساسیت بیشتر به مثال‌های آموزشی می‌شود، از تخمین توزیع پسین ماتریس فاصله استفاده شده است. در [19] به‌طور مشخص این روش بر روی الگوریتم معروف LMNN[5] [21] اعمال شده و باعث مقاوم‌پذیری بهتر آن در برابر نوفه برچسب شده است.

RDML[6] [22] برای مقاوم‌سازی الگوریتم یادگیری متریک، از تابع ضرر جدید و مقاوم Rescaled Hinge استفاده می‌کند. در این روش ابتدا مسأله یادگیری متریک با این تابع ضرر فرموله شده و سپس یک الگوریتم کارا بر اساس برنامه‌سازی نیمه درجه دوم[7]، برای حل آن ارائه‌شده است.

همان‌طور که مشاهده می‌شود کارهای فعلی به پیش‌پردازش داده‌ها برای مقاوم‌سازی الگوریتم متریک توجه ندارند؛ همچنین بیشتر آن‌ها بر روی تشخیص محدودیت‌های نوفه‌ای معطوف هستند، درحالی‌که تشخیص داده‌های حاوی نوفه برچسب، برای مقاوم‌پذیری الگوریتم یادگیری متریک بسیار مؤثرتر است.

در این مقاله یک تابع توزیع احتمال جدید معرفی شده است که هم‌زمان با تولید محدودیت‌های دوگانه یا سه‌گانه موردنیاز الگوریتم متریک، احتمال نوفه‌ای‌بودن برچسب داده‌ها را تقریب می‌زند. در ادامه الگوریتم یادگیری متریک جدیدی ارائه شده است که با وزن‌دهی سه‌گانه‌های آموزشی، تأثیر داده‌های حاوی نوفه برچسب را در فرایند یادگیری متریک به‌طور مؤثر کاهش می‌دهد و باعث مقاوم‌پذیری متریک یاد گرفته‌شده در برابر آن‌ها می شود.

از ویژگی‌های جالب روش ارائه‌شده، امکان اعمال ساده آن بر روی انواع مختلف الگوریتم‌های یادگیری متریک و مقاوم‌کردن آن‌ها در برابر نوفه برچسب است. در بخش بعد الگوریتم پیشنهادی با جزییات بیشتر بررسی می‌شود.

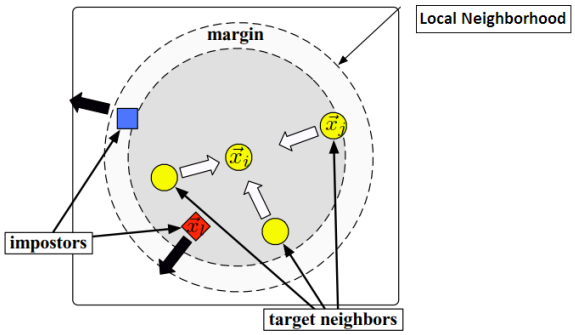
۲- روش پیشنهادی

همان‌طور که گفته شد، بیش‌تر روش‌های یادگیری متریک از محدودیت‌های زوج و یا سه‌گانه استفاده می‌کنند. روش‌های مبتنی بر محدودیت‌های سه‌گانه، به‌طورمعمول به دلیل تأکید بر فاصله نسبی به‌جای مطلق عملکرد

[5] Large Margin Nearest Neighbor
[6] Robust Distance Metric Learning
[7] Half Quadratic (HQ)

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

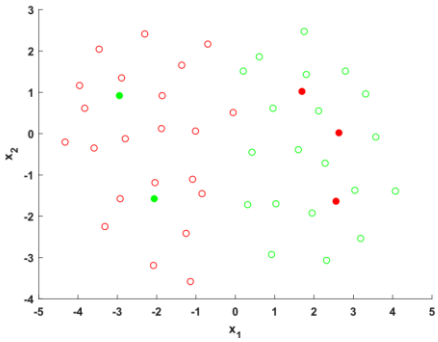
بالاتری دارند. به‌منظور تولید سه‌گانه‌های آموزشی در بیش‌تر الگوریتم‌ها یادگیری متریک، برای هر داده x_i k-نزدیک‌ترین همسایه که برچسب (کلاس) یکسان با آن دارند، به‌عنوان داده‌های مشابه اعلان می‌شوند. به این داده‌ها، همسایه‌های هدف x_i گفته می‌شود[1]. هم‌چنین به داده‌هایی که در محدوده فاصله x_i با دورترین همسایه هدف آن، با درنظرگرفتن یک حاشیه، قرار دارند، همسایه‌های بَدَل[2] x_i گفته می‌شود و x_i با آن‌ها نامشابه اعلان می‌شود. برای درک بهتر مفاهیم همسایه‌های هدف و بدل به شکل (۲) مراجعه شود. پس از ایجاد مجموعه‌های مشابه و نامشابه (S و D)، محدودیت‌های سه‌گانه T از پیوند طبیعی مجموعه‌های S و D تولید می‌شود. در روش پیشنهادی برای جلوگیری از تأثیر داده‌های پرت و نوفه برچسب، این محدودیت که همسایه‌های هدف و بدل یک داده باید متعلق به $2k$ همسایه نزدیک آن باشند، اضافه شده است.



(شکل-۲): همسایه‌های هدف و بدل داده x_i

(Figure-2): Target neighbors and imposters of x_i

برای درک بهتر مبنای روش پیشنهادی، در این بخش از یک مجموعه‌داده ساختگی حاوی نوفه برچسب که در شکل زیر مشاهده می‌شود، استفاده می‌کنیم.



(شکل-۳): مجموعه‌داده ساختگی حاوی نوفه برچسب

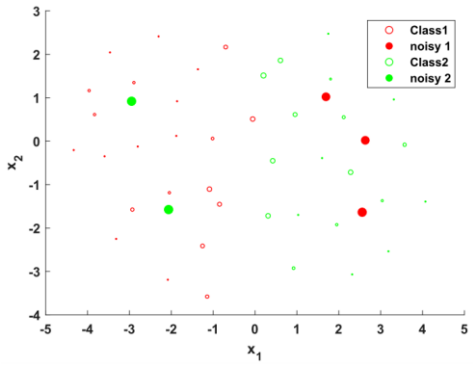
(Figure-3): A syntactic dataset containing label

در شکل (۳) داده‌های دارای نوفه برچسب به‌صورت دایره توپُر نشان داده شده است. الگوریتم گفته‌شده برای تولید سه‌گانه‌های آموزشی را با $k = 3$ بر روی این مجموعه‌داده اعمال می‌کنیم. با اعمال این الگوریتم، 306 سه‌گانه آموزشی تولید می‌شود که در 210 (معادل 68.63%) عدد از این سه‌گانه‌ها، نوفه برچسب وجود دارد. این امر نشان می‌دهد که حتی تعداد کمی داده با نوفه برچسب، تأثیر زیادی بر نوفه‌ای‌شدن سه‌گانه‌های آموزشی دارد. اکنون متغیرهای زیر را تعریف می‌کنیم:

- imp_i: تعداد دفعاتی که داده x_i به‌عنوان *همسایه بَدَل* در مجموعه سه‌گانه‌های آموزشی T معرفی شده است.
- tar_i: تعداد دفعاتی که داده x_i به‌عنوان *همسایه هدف* در مجموعه سه‌گانه‌های آموزشی T معرفی شده است

انتظار می‌رود چون مجموعه‌داده‌های دارای نوفه برچسب از سایر داده‌های رده مربوط دور هستند، به‌وسیلهٔ تعداد کمی از داده‌ها به‌عنوان همسایه مشابه معرفی شوند، از طرفی به دلیل اینکه این داده‌ها در میان داده‌های کلاس دیگر حضور دارند، توسط تعداد زیادی از آن‌ها به‌عنوان همسایه *بَدَل* معرفی شوند. ازاین‌رو، نسبت $\frac{imp_i}{(tar_i+1)}$ برای داده‌های حاوی نوفه برچسب بزرگ است و می‌توان با این معیار آن‌ها را شناسایی کرد. فرض کنید متغیر دودویی l_i نشان می‌دهد که داده x_i حاوی نوفه برچسب است یا خیر. با توجه به مطالب گفته‌شده توزیع این متغیر را به‌صورت زیر تعریف می‌کنیم:

$$\mathrm{p}(l_i = 1) = 1 - \exp\left(-\frac{imp_i/(tar_i + 1)}{\sigma^2}\right) \qquad (۲)$$



(شکل-۴): مجموعه‌داده ساختگی در شکل قبلی در آن اندازه هر داده متناسب با احتمال نوفه‌ای‌بودن برچسب آن $\mathrm{p}(l_i = 1)$ نشان داده‌شده است

(Figure-4): The syntactic dataset shown in the previous figure. Here, the size of each data point is proportional to its label noise probability ($\mathrm{p}(l_i = 1)$).

دقت کنید که برای داده‌های نوفه‌ای با توجه به‌دلیل بزرگ‌بودن $\frac{imp_i}{(tar_i+1)}$، $\exp\left(-\frac{imp_i/(tar_i+1)}{\sigma^2}\right) \to 0$ و درنتیجه $\mathrm{p}(l_i = 1) \to 1$. در شکل (۴) همان مجموعه‌داده ساختگی قبلی را مشاهده می‌کنید که در آن اندازه هر نقطه داده‌ای، احتمال نوفه‌ای‌بودن برچسب آن را نشان می‌دهد که با رابطه (۲) محاسبه شده است. در این

[1] Target Neighbors
[2] Imposter

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

آزمایش $\sigma^2 = 3$ در نظر گرفته شده است.

همان‌طور که مشاهده می‌شود، روش بالا با دقت قادر به شناسایی داده‌های حاوی نوفه برچسب است. نتایج آزمایش‌ها بر روی مجموعه‌داده‌های واقعی مختلف که در بخش بعد ارائه شده است، صحت این ادعا را به‌وضوح تأیید می‌کند. پس از تعیین احتمال نوفه‌ای‌بودن برچسب هر داده آموزشی، به مسأله یادگیری متریک می‌پردازیم.

فرض کنید مجموعه‌داده‌های آموزشی، سه‌گانه‌های T هستند. وزن هر سه‌گانه آموزشی $(\boldsymbol{x}_i, \boldsymbol{x}_i^+, \boldsymbol{x}_i^-)$ را برابر با $C_i = C\mathrm{p}(l_i = 0) = C\exp\left(-\frac{imp_i/(tar_i+1)}{\sigma^2}\right)$ تعریف می‌کنیم که C یک ضریب ثابت است. برای حل یادگیری متریک $\boldsymbol{M}$، مسأله بهینه‌سازی زیر پیشنهاد می‌شود:

$$\begin{aligned}&\underset{\boldsymbol{M}\succcurlyeq 0,\xi}{\text{minimize}}\ f(M,\boldsymbol{\xi}) = \frac{1}{2}\|\boldsymbol{M}\|_F^2 + \sum_{i=1}^{N} C_i\xi_i \\ &\text{subject to} \\ &1 - d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_i^-) + d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_i^+) \le \xi_i, \quad \xi_i \ge 0 \quad i = 1,2,\dots,N\end{aligned} \tag{۳}$$

در رابطه بالا $\|\boldsymbol{M}\|_F$ نشان‌دهنده نرم فروبینوس (Frobenius norm) ماتریس $\mathbf{M}$ است. بردار $\boldsymbol{\xi}$ نقش متغیر لغزش (slack variable) را دارد و اجازه می‌دهد که امکان خطا بر روی مثال‌های آموزشی وجود داشته باشد تا از بیش برازش (overfitting) جلوگیری به‌عمل آید. همانند SVM، در این رابطه از تابع ضرر مبتنی بر حاشیه Hinge استفاده شده است که بر روی سه‌گانه $(\boldsymbol{x}_i, \boldsymbol{x}_i^+, \boldsymbol{x}_i^-)$ به‌صورت زیر تعریف می‌شود:

$$l(x_i, x_i^+, x_i^-) = \max\{1 - d_M(x_i, x_i^-) + d_M(x_i, x_i^+), 0\} \tag{۴}$$

توجه داشته باشید که به‌دلیل شرط $\xi_i \ge 0$ در رابطه (۳)، نیازی به استفاده از max در مسأله بهینه‌سازی نیست.

به‌این‌ترتیب، به سه‌گانه‌های آموزشی مربوط به داده‌هایی که با احتمال زیاد دارای نوفه برچسب هستند وزن کمتری داده می‌شود و تأثیر آن‌ها در فرایند یادگیری کاهش می‌یابد. مسأله بالا مشابه مسأله بهینه‌سازی ارائه‌شده در [23] است با این تفاوت که در آن وزن سه‌گانه‌ها ثابت نیست. برای حل آن، ابتدا تابع لاگرانژ L را می‌نویسیم و پس‌ازآن نسبت به $\boldsymbol{M}$ و ξ_i مشتق گرفته و برابر با صفر قرار می‌دهیم. نتایج زیر به‌دست می‌آید:

$$\begin{aligned}L = &\frac{1}{2}\|\boldsymbol{M}\|_F^2 + \sum_{i=1}^{N} C_i\xi_i \\ &- \sum_{i=1}^{N} \alpha_i(1 - d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_i^-) + d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_i^+) - \xi_i) \\ &- \sum_{i=1}^{N} \beta_i\xi_i \\ &where\ \alpha_i, \beta_i \ge 0\end{aligned} \tag{۵}$$

که در رابطه بالا α_i و β_i ضرایب لاگرانژ هستند.

$$\frac{\delta L}{\delta \boldsymbol{M}} = 0 \Rightarrow \boldsymbol{M} = \sum_{i=1}^{N} \alpha_i \boldsymbol{A}_i \quad where \quad \boldsymbol{A}_i = (\boldsymbol{x}_i - \boldsymbol{x}_i^-)(\boldsymbol{x}_i - \boldsymbol{x}_i^-)^T - (\boldsymbol{x}_i - \boldsymbol{x}_i^+)(\boldsymbol{x}_i - \boldsymbol{x}_i^+)^T \tag{۶}$$

$$\frac{\delta L}{\delta \xi_i} = 0 \Rightarrow C_i - \alpha_i - \beta_i = 0 \Rightarrow 0 \le \alpha_i \le C_i \tag{۷}$$

با توجه به رابطه (۶)، برای تعیین متریک بهینه $\boldsymbol{M}$ نیاز به دانستن مقادیر α_i داریم. برای تعیین این ضرایب، باید مسأله دوگان را حل کنیم. با جایگزینی روابط (۶) و (۷) در (۵)، مسأله بهینه‌سازی دوگان به‌صورت زیر به‌دست می‌آید:

$$\begin{aligned}&\underset{\boldsymbol{\alpha}}{\text{maximize}} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j k(t_i, t_j) + \sum_{i=1}^{N} \alpha_i \\ &\text{subject to} \\ &0 \le \alpha_i \le C_i, \quad i = 1,2,\dots,N\end{aligned} \tag{۸}$$

که در رابطه بالا تابع کرنل $k(t_i, t_j)$ به‌صورت زیر تعریف شده است:

$$k(t_i, t_j) = trace(\boldsymbol{A}_i, \boldsymbol{A}_j) \tag{۹}$$

مسأله دوگان به‌سادگی توسط کتابخانه‌های آماده SVM مانند libSVM [۲۴] حل است. پس‌ازآن با استفاده از رابطه (۶)، مقدار بهینه ماتریس فاصله $\boldsymbol{M}$ به‌دست می‌آید. در الگوریتم (۱) مراحل روش پیشنهادی خلاصه شده است. دقت کنید که برای برقراری قید نیمه‌معین مثبت‌بودن ماتریس $\boldsymbol{M}$، کافی است الگوریتم تجزیه مقادیر ویژه (Eigen value) را بر روی آن اعمال کنیم و مقادیر ویژه نامنفی را مساوی صفر قرار دهیم. این عمل در الگوریتم (۱) توسط تابع $\mathrm{psd}(\boldsymbol{M})$ انجام می‌شود.

Algorithm1. RTripDML (Robust Triplet-based Distance Metric Learning)

Inputs:
$\boldsymbol{X}$: training instances
$\boldsymbol{T}$: set of training triplets

1.Compute imp_i and tar_i for each training instance $\boldsymbol{x}_i$
2. Compute $C_i = C\mathrm{p}(l_i = 0) = C\exp\left(-\frac{imp_i/tar_i}{\sigma^2}\right)$
3. Solve optimization problem (3) using weighted libsvm package
4. Obtain **M** using (6) 5. $\boldsymbol{M} = psd(\boldsymbol{M})$

دقت کنید که روش پیشنهادی برای تشخیص داده‌های حاوی نوفه برچسب و وزن دهی سه‌گانه‌ها وابسته به الگوریتم یادگیری متریک خاصی نیست و به‌سادگی می‌توان آن را بر روی الگوریتم‌های فعلی یادگیری متریک نیز اعمال کرد. همچنین از نتایج به‌دست‌آمده به‌وسیلهٔ روش پیشنهادی می‌توان برای بهبود الگوریتم مبتنی بر متریک، مانند kNN نیز بهره برد و آن را در مقابل نوفه

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

برچسب مقاوم کرد. ازاین‌رو در آزمایش‌ها، برای طبقه‌بندی داده‌ها توسط روش پیشنهادی، به‌جای الگوریتم kNN عادی از الگوریتم زیر که Robust_kNN نامیده شده است، استفاده کرده‌ایم.

Algorithm2. Robust_kNN

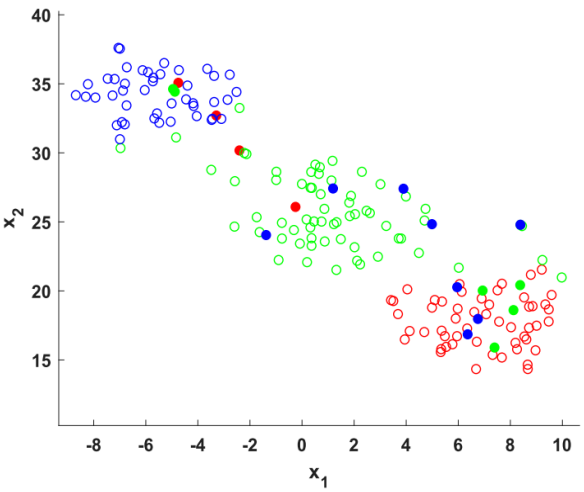
Inputs:

$\boldsymbol{X}$: training instances
$\boldsymbol{X}_{test}$: set of test examples
k: k parameter in kNN
$\boldsymbol{p}$: label noise probability vector
η: Percentage of noisy labels

1. Sort data according to $\boldsymbol{p}$ values
2. remove η percent of data in **X** with the highest label noise probability
3. **for each** instance $\boldsymbol{x}_{test}$ in $\boldsymbol{X}_{test}$
   3.1 Compute k nearest neighbors in Set $\boldsymbol{X}$
   3.2 Predict label of $\boldsymbol{x}_{test}$ based on the weighted majority voting of its neighbors. Here, we consider weight of each neighbor equals to $\mathrm{p}(l_i = 1)$

۳- آزمایش‌ها و تحلیل نتایج

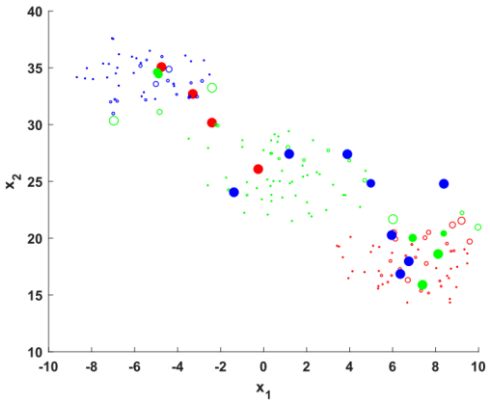
برای بررسی دقت روش پیشنهادی در تشخیص نوفه برچسب، در نخستین آزمایش، از مجموعه‌داده واقعی Wine از UCI استفاده‌شده است. ابتدا برچسب ۱۰٪ از داده‌ها را به‌طور تصادفی تغییر می‌دهیم. شکل زیر نتیجه اعمال نوفه برچسب بر روی این مجموعه‌داده را نشان می‌دهد. دقت کنید که به‌دلیل اینکه بُعد اصلی داده‌ها در این مجموعه‌داده برابر ۱۳ است، برای نمایش اطلاعات بُعد داده‌ها با استفاده از الگوریتم tSNE [25] به ۲ کاهش داده شده است.



(شکل-۵): مجموعه داده کاهش بُعد یافته Wine با اعمال 10% نوفه برچسب تصادفی

(Figure-5): The Wine dataset after dimensionality reduction and applying 10% label noise.

در ادامه با استفاده از الگوریتم یادشده، سه‌گانه‌ها را تولید می‌کنیم و احتمال نوفه‌ای‌بودن برچسب هر داده را با استفاده از رابطه (۲) محاسبه می‌کنیم. در شکل زیرمجموعه داده Wine را مشاهده می‌کنید که در آن اندازه هر نقطه داده‌ای، متناسب با احتمال نوفه‌ای‌بودن برچسب آن است. در این آزمایش $\sigma^2 = 5$ در نظر گرفته شده است.



(شکل-۶): مجموعه داده کاهش بعد یافته Wine با اعمال نوفه برچسب 10%. اندازه هر داده متناسب با احتمال نوفه‌ای‌بودن برچسب آن ($\mathrm{p}(l_i = 1)$) نشان داده شده است

(Figure-6): The Wine dataset after dimensionality reduction and applying 10% label noise. Here, the size of each data point is proportional to its label noise probability ($p(l_i = 1)$).

نتایج این آزمایش تأیید می‌کند که روش پیشنهادی در مجموعه‌داده واقعی Wine، همانند مجموعه‌داده ساختگی بخش قبلی (به شکل (۴) مراجعه شود) به‌درستی قادر به شناسایی داده‌ها با برچسب نوفه‌ای است.

برای تأیید بیشتر کارایی روش پیشنهادی، در ادامه این روش را بر روی انواع مختلف مجموعه‌داده‌های واقعی از کاربردهای مختلف در سطوح مختلف نوفه برچسب، آزمایش کرده و نتایج حاصل را با روش‌های همتا در مرزهای دانش مقایسه می‌کنیم. ویژگی‌های دادگان مورد آزمایش در جدول زیر خلاصه شده است.

(جدول-۱): ویژگی‌های مجموعه‌داده‌های مورداستفاده در آزمایش‌ها

(Table-1): Statistics of evaluated datasets

Data Set	# classes	# samples	# dim
Wine	3	178	13
Letters	26	20000	16
Extended Yale Faces	38	2414	4096
Ionosphere	2	351	34
Caltech101	101	9143	10000
WDBC	2	659	30
Australian	2	690	14
German Credit	2	1000	24

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

(جدول-۲): دقت طبقه‌بند kNN با استفاده از متریک یاد گرفته‌شده توسط روش پیشنهادی و سایر روش‌ها برحسب نوفه برچسب (nl)

(Table-2): The kNN classifier accuracy using the learned metric of the proposed and other competing methods.

Data Set	nl %	RTripDML	RDML	BLMNN	L1_DML	SVM_TRIP	Euc
Wine	0	96.98±0.97	**97.55+-2.19**	96.60+-1.95	96.23+-1.78	97.36+-1.32	94.91+-2.36
	5	**96.79+-3.09**	96.23+-2.81	95.66+-2.82	95.47+-0.97	94.91+-3.09	94.15+-2.43
	10	**96.98+-2.39**	96.57+-3.73	94.53+-3.61	94.72+-3.54	92.45+-4.17	92.45+-3.88
	15	**94.91+-2.52**	93.71+-3.13	90.94+-4.86	90.94+-4.34	90.19+-5.40	90.19+-3.65
	20	**94.91+-3.33***	91.32+-4.72	91.32+-3.47	86.98+-3.92	89.06+-2.64	87.55+-2.70
Ionosphere	0	90.86+-2.96	90.86+-4.12	86.29+-3.59	88.57+-3.19	**91.43+-3.64**	83.71+-3.73
	5	**89.14+-2.26**	87.14+-5.98	85.14+-5.01	83.71+-3.99	88.00+-3.73	83.71+-3.29
	10	**89.14+-2.96***	84.57+-3.10	84.00+-3.83	81.14+-4.89	84.57+-4.78	80.00+-2.86
	15	**89.71+-2.93***	84.00+-2.12	83.71+-5.31	82.29+-4.58	86.00+-2.35	80.57+-3.73
	20	**88.86+-3.70***	82.57+-4.89	82.00+-3.86	79.14+-8.72	80.57+-3.86	78.86+-3.41
WDBC	0	**97.70+-1.34**	**97.70+-1.73**	94.51+-1.31	93.10+-0.74	95.75+-1.15	92.74+-1.45
	5	**96.64+-2.02**	96.11+-1.48	94.16+-1.61	92.04+-1.77	93.45+-1.73	91.86+-2.29
	10	**96.11+-1.48**	94.16+-1.61	90.97+-1.15	89.20+-2.46	90.27+-3.54	89.73+-2.63
	15	**95.22+-1.34***	89.56+-2.11	88.50+-2.08	85.49+-4.18	88.32+-2.20	86.73+-3.48
	20	**93.63+-1.70***	83.89+-5.92	84.42+-3.10	82.48+-1.92	83.72+-2.55	83.36+-3.09
Letters	0	96.48+-0.20	**97.44±0.17**	95.91+-0.24	97.20+-0.25	95.62+-0.29	95.07+-0.39
	5	95.11+-0.19	**96.47+-0.38***	94.51+-0.34	95.05+-0.53	94.49+-0.17	93.98+-0.29
	10	**94.70+-0.22**	94.55+-0.34	92.58+-0.28	91.97+-0.44	93.05+-0.19	92.15+-0.25
	15	**94.45+-0.33***	91.77+-0.44	90.10+-0.34	89.39+-0.49	91.01+-0.22	89.79+-0.35
	20	**93.90+-0.38***	88.42+-0.17	87.13+-0.13	86.30+-0.21	88.01+-0.30	86.42+-0.55
Extended Yale Faces	0	93.78+-1.57	**96.72±0.84***	91.70+-1.37	92.66+-0.67	**92.86+-1.07**	89.38+-1.23
	5	93.24+-1.36	**95.72±0.59***	90.41+-1.34	90.83+-0.87	90.17+-1.68	88.34+-1.19
	10	91.33+-2.18	**93.94±1.00**	89.21+-0.89	88.30+-0.81	87.30+-1.53	86.80+-0.81
	15	**90.50+-1.42**	89.83+-0.96	86.51+-1.28	84.48+-1.83	84.44+-1.91	84.56+-1.07
	20	**89.38+-1.63**	89.21±0.82	82.82+-2.43	80.75+-2.50	80.79+-1.76	81.33+-1.45
Australian	0		86.52+-2.02				
	5	**87.25+-4.88**	84.20+-3.13	84.64+-4.14	85.22+-2.79	84.64+-4.14	84.06+-3.97
	10	**86.38+-3.49**	84.20+-3.81	84.06+-5.12	83.62+-2.09	84.06+-5.12	83.48+-3.01
	15	**84.20+-2.83**	82.61+-3.59	82.46+-3.05	83.04+-2.21	82.75+-3.45	83.33+-4.99
	20	**84.35+-3.18**	77.10+-5.56	80.00+-3.46	80.43+-3.28	79.42+-4.60	81.74+-4.51
		81.45+-4.71		75.51+-3.49	73.91+-2.71	74.78+-2.96	77.10+-5.56
German Credit	0		73.30+-1.20				
	5	**72.40+-3.78**	71.80+-5.02	70.40+-5.19	70.30+-5.46	70.90+-2.79	69.60+-1.34
	10	**71.80+-3.37**	70.00+-3.00	70.20+-3.95	68.20+-2.31	69.80+-1.79	67.80+-3.01
	15	**70.30+-4.32**	67.40+-4.25	67.40+-3.65	68.40+-3.36	67.20+-2.80	66.70+-2.39
	20	**69.50+-3.06**	67.60+-4.16	66.10+-4.20	67.60+-4.67	67.90+-5.02	65.70+-3.77
		69.50+-3.79		65.00+-4.32	66.80+-4.42	66.60+-4.57	65.20+-4.10
Caltech101-10	0		97.59+-0.38				
	5	**97.80+-1.15**	96.99+-0.50	96.40+-1.19	97.20+-1.30	97.30+-1.25	95.90+-0.55
	10	**97.70+-1.48**	96.45+-0.49	96.30+-1.20	97.30+-0.76	96.40+-1.56	95.60+-0.74
	15	**97.30+-1.04**	95.60+-0.75	95.40+-0.82	96.40+-0.42	95.80+-1.48	94.60+-0.96
	20	**97.60+-1.64***	93.37+-0.84	94.50+-1.94	94.90+-1.14	95.40+-1.75	93.00+-1.54
		97.40+-1.52*		92.60+-2.27	93.10+-1.34	93.70+-2.28	91.40+-1.29
متوسط رتبه Avg. Rank	0	**1.87**	1.62	4.50	4.00	3.00	6.00
	5	**1.25**	2.00	3.62	4.12	4.12	5.87
	10	**1.25**	1.87	4.12	4.25	4.25	5.37
	15	**1.00**	2.37	4.00	4.62	3.87	5.12
	20	**1.00**	2.25	3.75	5.25	3.87	4.87
Total Avg Rank		**1.25**	2.02	4.00	4.45	3.82	5.45

در مجموعه‌داده‌های مورد آزمایش، به‌دلیل اینکه نرمال‌کردن ویژگی‌ها باعث بهبود نتایج برای همه روش‌ها می‌شود، میانگین هر ویژگی برابر صفر و واریانس آن برابر یک تنظیم شده است. در مجموعه‌داده Extended Yale Faces برای حذف ویژگی‌های نوفه‌ای، بُعد تصاویر با استفاده از PCA به صد کاهش داده شده است. در مجموعه‌داده Caltech101 برای استخراج ویژگی از شبکه عصبی پیچشی عمیق imagenet-resnet-152-dag[1] استفاده‌شده است و برای هر تصویر، هزار ویژگی از لایه ماقبل آخر این شبکه استخراج‌شده است.

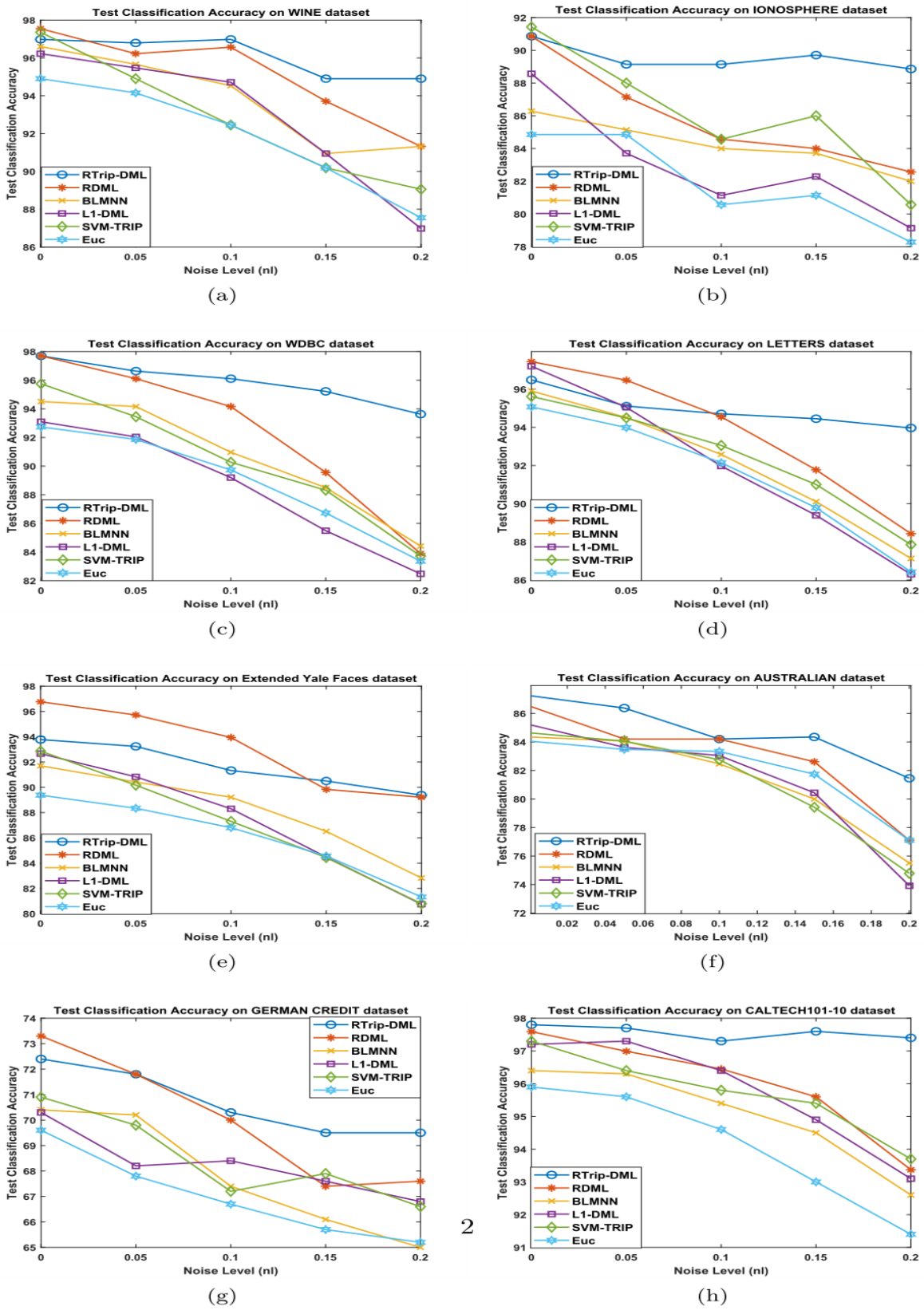
روش‌های رقیب در این آزمایش‌ها RDML [22]، BLMNN [19]، L1-DML [17]، SVM-Trip [23] و روش Euc (اقلیدسی) انتخاب شده‌اند که کد منبع آن‌ها در شبکه اینترنت قابل‌دسترس است.

در جدول (۲) دقت دسته‌بندی روش‌های مختلف باهم مقایسه شده است. نتایج ارائه‌شده، میانگین دقت طبقه‌بند kNN به‌همراه متریک آموزش‌دیده‌شده به‌وسیلهٔ روش‌های مختلف با اعتبارسنجی متقاطع $k = 5$ را نشان می‌دهد. در روش پیشنهادی به‌جای الگوریتم kNN از الگوریتم Robust_kNN استفاده‌شده است.

[1] https://www.vlfeat.org/matconvnet/pretrained/

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]



(شکل-۷): مقایسه دقت طبقه‌بندی روش پیشنهادی RTripDML با روش‌های رقیب برحسب درصد نوفه برچسب

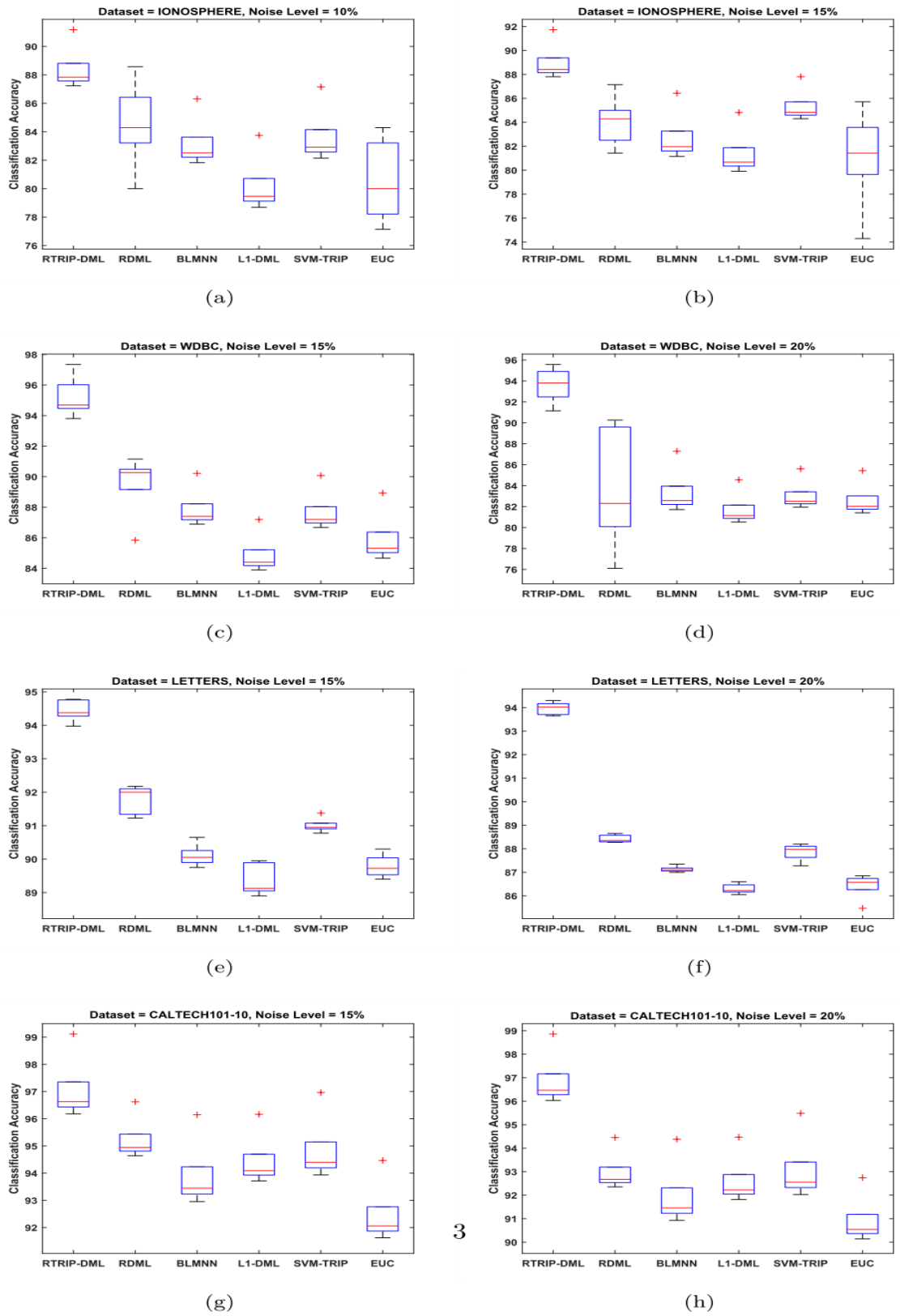
(Figure-7): Comparison of classification accuracy of the proposed and other competing methods vs label noise

نوفه برچسب، پیچیده‌ترین مدل نوفه است که برای تولید آن، درصد مشخصی از داده‌ها به‌طور تصادفی انتخاب شده‌اند و برچسب آن‌ها به‌صورت تصادفی به سایر رده‌ها تغییر یافته است. پارامتر nl در این جدول، نشان‌دهنده درصد نوفه برچسب است.

برای مثال $nl = 20$ نشان می‌دهد که برچسب ۲۰٪ داده‌ها به‌صورت تصادفی تغییر پیدا کرده است. برای درک بهتر نتایج، در شکل (۷) دقت روش‌های مختلف در سطوح مختلف نوفه برچسب نشان داده شده است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]



(شکل-۸): نمودار جعبه برخی نتایج که ازنظر آزمون آماری با $p\text{-}value = 5\%$ دارای اختلاف معنادار هستند

(Figure-8): Box-Plots of some results with the statistically significant differences ($p\text{-}value = 5\%$).

مقدار ابر پارامترها با اعتبارسنجی متقاطع[1] با $k = 5$ تنظیم شده است. به بیان دقیق‌تر، مقدار C در آزمایش‌ها از بازه $(10^{-6}, 1)$ انتخاب شده است. پارامتر λ در روش‌های BLMNN و L1_DML از بازه (۱و۲۰) و مقدار η در روش پیشنهادی RTrip-DML از محدوده (۱و۵) تنظیم شده است.

[1] Cross Validation

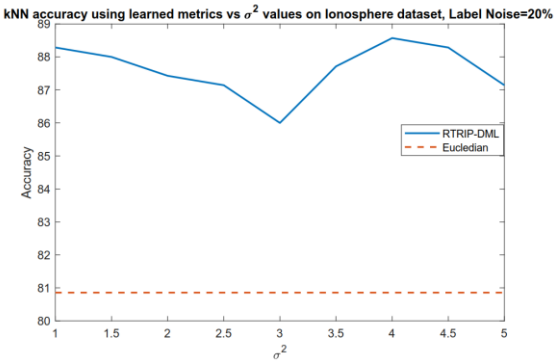
[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]

برای درک بهتر نتایج و مقایسه روش‌ها، آزمون آماری با $p\text{-}value = 5\%$ بر روی نتایج انجام شده است و نتایجی که ازنظر این آزمون اختلاف معنادار با سایر نتایج را دارند، با * در این جدول علامت‌گذاری شده‌اند. همچنین، میانگین رتبه کلی و رتبه به تفکیک سطح نوفه برای هر الگوریتم مشخص‌شده و به این جدول اضافه‌شده است؛ به‌علاوه برای درک بهتر نتایج، در شکل (۸) نمودار جعبه (Box-Plot) برخی از نتایج که ازنظر آماری دارای اختلاف معنادار هستند، رسم شده است.

همان‌طور که جدول (۲) و شکل (۷) نشان می‌دهد، روش پیشنهادی نسبت به سایر روش‌های قابل‌مقایسه در تمام مجموعه‌داده‌های مورد آزمایش برتری قابل‌ملاحظه دارد و کارایی آن با افزایش درصد داده‌های حاوی نوفه برچسب، افت به‌مراتب کمتری دارد. همچنین در سطوح بالای نوفه برچسب سایر روش‌های یادگیری متریک عملکردی مشابه روش اقلیدسی دارند و در این محیط‌ها یادگیری متریک چندان مؤثر نیست.

نتایج این آزمایش‌ها تأیید می‌کند که روش پیشنهادی قادر به تشخیص مؤثر داده‌های پرت و داده‌های حاوی نوفه برچسب در مجموعه‌داده‌های مختلف است و تأثیر آن‌ها را در فرآیند آموزش و آزمون به‌طور قابل ملاحظه‌ای کاهش می‌دهد.



(شکل-۹): بررسی تأثیر پارامتر σ^2 در مقاوم‌پذیری روش پیشنهادی در مجموعه‌داده Ionosphere با نوفه برچسب ۲۰٪

(Figure-9): The effect of parameter σ^2 in the accuracy of the proposed method in the Ionosphere dataset contaminated with 20% label noise.

برای بررسی تأثیر پارامتر η در مقاوم‌پذیری الگوریتم پیشنهادی، در آزمایش دیگر در دادگان Ionosphere با نوفه ۲۰٪، نتایج روش پیشنهادی Robust-ODML با روش اقلیدسی مقایسه شده است. مقدار بهینه پارامتر C در این آزمایش با استفاده از اعتبارسنجی متقاطع برابر با $C = .5$ تعیین شده است. شکل (۹) نتایج حاصل از این آزمایش را نشان می‌دهد.

همان‌طورکه مشاهده می‌شود، روش پیشنهادی در این مجموعه‌داده حاوی نوفه برچسب به‌ازای تمامی مقادیر σ^2 عملکرد به‌مراتب بهتری نسبت به روش اقلیدسی دارد و میزان حساسیت سامانه پیشنهادی به مقدار این پارامتر به‌نسبه پایین است.

۴- جمع‌بندی و نتیجه‌گیری

با افزایش روزافزون حجم داده‌ها و هزینه‌های بالای برچسب‌زنی آن‌ها، امروزه برای تعیین برچسب داده‌ها به‌طورمعمول از اینترنت و با روش‌هایی مانند Crowdsourcing استفاده می‌شود. این اطلاعات می‌تواند حاوی نوفه و برچسب اشتباه باشد. روش‌های عادی یادگیری متریک در مقابل نوفه برچسب مقاوم نیستند و کارایی آن‌ها در محیط‌های حاوی نوفه برچسب اُفت قابل‌ملاحظه‌ای می‌کند. این امر منجر به ارائه روش‌های یادگیری متریک مقاوم شده است.

در این پژوهش یک روش احتمالاتی مؤثر برای شناسایی داده‌های حاوی نوفه برچسب ارائه شده که با استفاده از محدودیت‌های مسأله یادگیری متریک، قادر است، احتمال نوفه‌ای‌بودن داده‌ها را تقریب بزند. در ادامه یک الگوریتم یادگیری متریک جدید ارائه شده است که با وزن‌دهی سه‌گانه‌های آموزشی، تأثیر داده‌های حاوی نوفه برچسب را در فرایند آموزش متریک، به‌طور مؤثر کاهش می‌دهد و باعث مقاوم‌پذیری متریک یادگرفته‌شده در برابر آن‌ها می‌گردد؛ همچنین این پژوهش نشان می‌دهد که چگونه از نتایج به‌دست‌آمده به‌وسیلهٔ روش پیشنهادی می‌توان برای بهبود الگوریتم مبتنی بر متریک مانند kNN نیز بهره برد و آن را در مقابل نوفه برچسب مقاوم کرد.

نتایج آزمایش‌ها بر روی مجموعه‌داده‌های مختلف تأیید می‌کند که روش پیشنهادی به‌خوبی قادر به شناسایی داده‌های حاوی نوفه برچسب و کاهش تأثیر آن‌ها در فرآیند یادگیری متریک و آزمون است. همچنین الگوریتم ارائه‌شده نسبت به روش‌های همتا مقاومت به‌مراتب بهتری در برابر نوفه برچسب دارد و کارایی آن حتی در محیط‌هایی با درصد نوفه برچسب بالا، اُفت بسیار کمی پیدا می‌کند.

برخی توصیه‌های برای کارهای آینده در این زمینه عبارتند از:

۱- بررسی کارایی روش پیشنهادی در سامانه‌های CBIR[1]

۲- توسعه روش پیشنهادی برای یادگیری نیمه‌ناظر[2]

5- References ۵- مراجع

[1] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian similarity learning based on posterior distribution of data," *Engineering Applications of Artificial Intelligence*, vol. 67,

[1] Content Based Information Retrieval
[2] Semi Supervised Learning

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]
[DOI: 10.52547/jsdp.19.1.125]

in the presence of label noise: a survey," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 5, pp. 845-869, 2013.

[13] T. Yang, R. Jin, and A. K. Jain, "Learning from noisy side information by generalized maximum entropy model," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010: Citeseer, pp. 1199-1206 .

[14] K. Huang, R. Jin, Z. Xu, and C.-L. Liu, "Robust metric learning by smooth optimization," *arXiv preprint arXiv:1203.3461*, 2012.

[15] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127-152, 2005.

[16] D. Wang and X. Tan, "Robust Distance Metric Learning in the Presence of Label Noise," in *AAAI*, 2014, pp. 1321-1327 .

[17] H. Wang, F. Nie, and H. Huang, "Robust Distance Metric Learning via Simultaneous L1-Norm Minimization and Maximization," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, T. Jebara and E .P. Xing, Eds., 2014, [Formatter not found: ResolvePDF]: JMLR Workshop and Conference Proceedings, pp. 1836-1844. [Online]. Available: http://jml-r.org/proceedings/papers/v32/wangj14.pdf. [Online]. Available: http://jmlr.org/proceedings/papers/v32/wangj14.pdf

[18] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recogn.*, vol. 41, no. 12, pp. 3600-3612, 2008, doi: 10.1016/j.patcog.2008.05.018.

[19] D. Wang and X. Tan, "Robust Distance Metric Learning via Bayesian Inference," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1542-1553, 2018.

[20] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian approach for metric learning in latent space," *Knowledge-Based Systems* , vol. 178, pp. 11-24, 2019.

[21] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207-244, 2009.

[22] S. Al-Obaidi, D. Zabihzadeh, A. S. Rasheed, and R. Monsefi" ,Robust Metric Learning based on the Rescaled Hinge Loss," *arXiv preprint arXiv:1904.11711*, 2019.

[23] F. Wang, W. Zuo, L. Zhang, D. Meng, and D. pp. 173-186, 2018.

[2] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang, "Cross-domain visual matching via generalized similarity measure and feature learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1089-1102, 2017.

[3] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 2, pp. 331-345, 2014.

[4] S. Bak and P. Carr, "One-Shot Metric Learning for Person Re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2990-2999 .

[5] N. Jiang, W. Liu, and Y. Wu, "Order determination and sparsity-regularized metric learning adaptive visual tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012: IEEE, pp. 1956-1963 .

[6] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009: IEEE, pp. 309-316 .

[7] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large Scale Online Learning of Image Similarity Through Ranking," *J. Mach. Learn. Res.*, vol. 11, pp. 1109-1135, 2010.

[8] X. Hao, S. C. H. Hoi, J. Rong, and Z. Peilin, "Online Multiple Kernel Similarity Learning for Visual Search," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 3, pp. 536-549, 2014, doi: 10.1109/TPAMI-.2013.149.

[9] P. Wu, S. C. H. Hoi, P. Zhao, C. Miao, and Z. Y. Liu, "Online Multi-Modal Distance Metric Learning with Application to Image Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 454-467, 2016, doi: 10.1109/TKDE.2015.2477296.

[10] J. Li, C. Xu, W. Yang, C. Sun, and D. Tao, "Discriminative Multi-View Interactive Image Re-Ranking," *IEEE Transactions on Image Processing*, 2017.

[11] A. Bellet, A. Habrard, and M. Sebban, "A Survey on Metric Learning for Feature Vectors and Structured Data," *Technical report*, 2014.

[12] B. Frénay and M. Verleysen, "Classification

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]

Zhang, "A kernel classification framework for metric learning," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 9, pp. 1950-1962, 2015.

[24] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[25] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579-2605, 2008.

داود ذبیح‌زاده فارغ‌التحصیل دکترای هوش مصنوعی از دانشگاه فردوسی مشهد است. زمینه‌های تخصصی مورد علاقه وی عبارتند از یادگیری ماشین، شناسایی آماری الگو، داده‌کاوی و شبکه های عصبی عمیق. در حال حاضر به‌عنوان استادیار گروه مهندسی کامپیوتر، دانشکده فناوری‌های نوین، دانشگاه حکیم سبزواری مشغول فعالیت است.

نشانی رایانامه ایشان عبارت است از:

d.zabihzadeh@gmail.com,
d.zabihzadeh@hsu.ac.ir

سعید زاهدی دانشجوی دکترای هوش مصنوعی از دانشگاه فردوسی مشهد است. از زمینه های تخصصی مورد علاقه اینجانب می‌توان به طراحی، تست و مدل‌سازی نرم‌افزار، یادگیری ماشین، شناسایی الگو و رباتیک اشاره کرد. در حال حاضر به‌عنوان مدیر کل فناوری اطلاعات استانداری گیلان مشغول فعالیت است.

نشانی رایانامه ایشان عبارت است از:

s.zahedi@mail.um.ac.ir

رضا منصفی کارشناسی برق و الکترونیک خود را از دانشگاه منچستر انگلستان در سال ۱۹۷۹ میلادی و کارشناسی ارشد و دکترای خود را از دانشگاه سالفورد انگلستان در مهندسی کنترل و در ارتباطات داده‌ها و کنترل نظارت به‌ترتیب در سال‌های ۱۹۸۱ و ۱۹۸۳ دریافت کرد. در شرکت رمزدک در شمال ویلز در سال‌های ۱۹۸۳ الی ۱۹۸۴ به‌عنوان مهندس ارشد پژوهش و توسعه فعالیت داشته و به مدت پنج سال از ۱۹۸۵ الی ۱۹۹۰ در دانشگاه میلتون کینز در کشور انگلستان به‌عنوان عضو هیئت علمی و عضو پژوهش‌گر (Research Fellow) همکاری داشته است. وی در سال ۱۳۷۰ شمسی به ایران بازگشت و از آن موقع تاکنون عضو هیئت علمی دانشگاه فردوسی مشهد است و در حال حاضر با مرتبه استادتمام با این دانشگاه همکاری می‌کند. شبکه‌های رایانه‌ای، شبکه‌های حس‌گر بی‌سیم و پردازش سیگنال و یادگیری ماشین در هوش مصنوعی از جمله علاقه‌های تخصصی وی است.

نشانی رایانامه ایشان عبارت است از:

monsefi@um.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-08-24]

[DOI: 10.52547/jsdp.19.1.125]