DOI: 10.22034/tjee.2022.15428

Semi-supervised Sparse Feature Selection based on Hessian Regularization and Fisher Discriminant Analysis

Razieh Sheikhpour

Department of Computer Engineering, Faculty of Engineering, Ardakan University, P.O. Box 184, Ardakan, Iran
E-mail: rsheikhpour@ardakan.ac.ir

Short Abstract
Feature selection is one of the most important techniques in machine learning and pattern recognition, which eliminates redudant features and selects a suitable subset of features. This avoids overfitting when building the model and improves the model performance. In many applications, obtaining labeled data is costly and time consuming, while unlabeled data are readily available. Therefore, semi-supervised feature selection methods can be used to consider both labeled and unlabeled data in the feature selection process. In this paper, a semi-supervised sparse feature selection method is proposed based on hessian regularization and Fisher discriminant analysis which selects the appropriate features using the labeled data and the local structure of both labeled and unlabeled data. In the proposed method, an objective function based on semi-supervised scatter matrix and $l_{2,1}$-norm is presented for feature selection which considers the correlation among features. To solve the proposed objective function, an iterative algorithm is used and its convergence is experimentally and theoretically proved. The results of the experiments on five data sets indicate that the proposed method improves the selection of relevant features compared to other methods used in this paper.

Keywords
Semi-supervised feature selection; Sparse models; Hessian regularization; Fisher discriminant analysis, $l_{2,1}$-norm.

1. Short Introduction

Feature selection is one of the most important techniques in machine learning and pattern recognition, which eliminates redudant features and selects a suitable subset of features. This avoids overfitting when building the model and improves the model performance. In many applications, obtaining labeled data is costly and time consuming, while unlabeled data are readily available. Therefore, semi-supervised feature selection methods can be used to consider both labeled and unlabeled data in the feature selection process. Classical semi-supervised feature selection methods evaluate the features one by one and neglect the correlation among different features. To overcome this problem, semi-supervised sparse feature selection methods have been proposed to consider the correlation among different featurs in feature selection process.

2. Proposed Work and Methodology

In this paper, a semi-supervised sparse feature selection method is proposed based on hessian regularization and Fisher discriminant analysis which selects the appropriate features using the labeled data and the local structure of both labeled and unlabeled data. In the proposed method, an objective function based on semi-supervised scatter matrix and l2,1-norm is presented for feature selection which considers the correlation among features. To solve the proposed objective function, an iterative algorithm is used and its convergence is experimentally and theoretically proved. The results of the experiments on five data sets indicate indicate that the proposed method outperforms other methods in selecting the relevant features.

3. Conclusion

In this paper, a semi-supervised sparse feature selection method based on hessian regularization and Fisher discriminant analysis was proposed which uses the labeled data and the local structure of both labeled and unlabeled data to select the appropriate features. Proposed method uses a semi-supervised scatter matrix and a regularization based on l2,1-norm to consider the correlation among features. An iterative algorithm is used to solve the proposed method and its convergence is experimentally and theoretically proved. The results of of the experiments showed that the proposed method improves the selection of relevant features compared to other methods used in this paper.

4. References

Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, S. Member, Semisupervised feature selection via spline regression for video semantic recognition, IEEE Trans. NEURAL NETWORKS Learn. Syst. 26 (2015) 252–264.
C. Shi, Q. Ruan, S. Member, G. An, R. Zhao, Hessian semi-supervised sparse feature selection based on L21/2-matrix norm, IEEE Trans. Multimed. 17 (2015) 16–28.

انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر

راضیه شیخ‌پور، استادیار
استادیار- گروه مهندسی کامپیوتر- دانشکده فنی و مهندسی- دانشگاه اردکان- اردکان- ایران-rsheikhpour@ardakan.ac.ir

چکیده

انتخاب ویژگی یکی از تکنیک‌های مهم در یادگیری ماشین و شناسایی الگو است که با حذف ویژگی‌های نامناسب و انتخاب زیرمجموعه‌ای مفید از ویژگی‌ها باعث اجتناب از بیش‌برازش در هنگام ساخت مدل، بهبود کارایی و سادگی مدل می‌شود. در بسیاری از کاربردها، تعیین برچسب داده‌ها هزینه‌بر بوده و مستلزم صرف زمان زیادی است، درحالی‌که داده‌های بدون برچسب به آسانی در دسترس هستند. بنابراین، استفاده از روش‌های انتخاب ویژگی نیمه‌نظارتی که بتوانند در فرآیند انتخاب ویژگی از داده‌های برچسب‌دار و بدون برچسب استفاده نمایند، بسیار ارزشمند است. در این مقاله، یک روش انتخاب ویژگی تُنک نیمه‌نظارتی مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر پیشنهاد می‌شود که می‌تواند با استفاده از داده‌های برچسب‌دار و اطلاعات توزیع و ساختار محلی داده‌های برچسب‌دار و بدون برچسب مناسب‌ترین ویژگی‌ها را انتخاب نماید. در روش پیشنهادی، تابع هدفی مبتنی بر ماتریس پراکندگی نیمه‌نظارتی و نُرم-$l_{2,1}$ برای انتخاب ویژگی ارائه می‌شود که از منظم‌سازی هسین و آنالیز تشخیصی فیشر در ساخت ماتریس پراکندگی نیمه‌نظارتی استفاده می‌کند و همبستگی بین ویژگی‌ها را در هنگام انتخاب ویژگی در نظر می‌گیرد. برای حل تابع هدف پیشنهادی مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر، الگوریتمی موثر با رویکرد تکراری به کار می‌رود و همگرایی آن به صورت تئوری و عملی اثبات می‌شود. نتایج به‌دست آمده از آزمایش‌ها بر روی پنج مجموعه داده حاکی از برتری روش پیشنهادی در مقایسه با دیگر روش‌های انتخاب ویژگی استفاده شده در این مقاله است.

کلمات کلیدی

انتخاب ویژگی نیمه‌نظارتی، مدل‌های تُنک، منظم‌سازی هسین، آنالیز تشخیصی فیشر، نُرم-$l_{2,1}$.

نام نویسنده مسئول: دکتر راضیه شیخ‌پور
ایمیل نویسنده مسئول: rsheikhpour@ardakan.ac.ir
تاریخ ارسال مقاله: ۱۴۰۰/۱۰/۱۳
تاریخ(های) اصلاح مقاله: ۱۴۰۱/۰۵/۰۷
تاریخ پذیرش مقاله: ۱۴۰۱/۰۶/۱۹

۱- مقدمه

پیشرفت‌های به وجود آمده در جمع‌آوری داده و قابلیت‌های ذخیره‌سازی در طی دهه‌های اخیر باعث شده است که در بسیاری از علوم، حجم بزرگی از داده‌ها به وجود آید که شامل تعداد زیاد نمونه‌ها و تعداد زیاد ویژگی‌های مربوط به یک نمونه است [۱، ۲]. داده‌هایی که دارای ویژگی‌های زیادی هستند، چالش‌های محاسباتی زیادی را ایجاد می‌کنند که یکی از مشکلات این است که در بیشتر مواقع، تمام ویژگی‌ها برای یافتن دانش نهفته در داده‌ها مهم و حیاتی نیستند [۲–۴]. به همین دلیل، در بسیاری از زمینه‌ها انتخاب ویژگی‌ها یا کاهش ابعاد داده‌ها یکی از مباحث قابل توجه است. روش‌های انتخاب ویژگی با حذف ویژگی‌های نامناسب و انتخاب زیرمجموعه‌ای مفید از ویژگی‌های اولیه، ابعاد داده‌ها را کاهش می‌دهند [۵، ۶]. انتخاب ویژگی باعث اجتناب از بیش‌برازش (Overfitting) در هنگام ساخت مدل، بهبود کارایی و سادگی مدل می شود. علاوه بر این، انتخاب ویژگی نقش مهمی در درک و تجسم داده‌ها ایفا می‌کند [۷، ۸].

روش‌های انتخاب ویژگی به سه دسته فیلتر (Filter)، یادگیرمبنا (Wrapper) و یادگیرحاصل (Embedded) تقسیم می‌شوند. روش‌های فیلتر مستقل از الگوریتم یادگیری، ویژگی‌ها را براساس یک مرحله پیش پردازش انتخاب می‌کنند. این روش‌ها از نظر محاسباتی خیلی ساده و سریع هستند و به راحتی برای مجموعه داده‌های با ابعاد بالا به کار می‌روند. روش‌های یادگیرمبنا، زیرمجموعه‌ای از ویژگی‌ها را براساس یک الگوریتم یادگیری انتخاب می‌کنند که باعث بیشینه شدن کارایی آن روش یادگیری شود. این روش‌ها کارایی قابل قبولی دارند ولی در عمل زمان‌گیر هستند. روش‌های فیلتر در مقایسه با روش‌های یادگیرمبنا سریع‌تر هستند. روش‌های یادگیرحاصل، انتخاب ویژگی و یادگیری مدل را به صورت هم‌زمان انجام می‌دهند که این عمل معمولاً با استفاده از یک تابع هدف انجام می‌شود [۹–۱۱].

علاوه بر تقسیم‌بندی فوق، روش‌های انتخاب ویژگی با توجه به برچسب کلاس‌ها به سه دسته نظارتی، غیرنظارتی و نیمه‌نظارتی تقسیم می‌شوند [۱۲]. روش‌های انتخاب ویژگی نظارتی، اهمیت ویژگی‌ها را با توجه به همبستگی ویژگی‌ها و برچسب داده‌ها ارزیابی می‌کنند [۱۳]. روش‌های انتخاب ویژگی غیرنظارتی، اهمیت ویژگی‌ها را با توجه به توانایی حفظ خصوصیات خاصی از داده‌ها مانند واریانس یا توانایی حفظ ساختار محلی ارزیابی می‌کنند [۱۴]. هنگامی‌که داده‌های برچسب‌دار کافی وجود داشته باشند، روش‌های انتخاب ویژگی نظارتی در مقایسه با روش‌های انتخاب ویژگی غیرنظارتی کارایی بالاتری دارند. در بسیاری از کاربردها، داده‌های برچسب‌دار با صرف زمان و هزینه زیادی بدست می‌آیند و تعداد کافی از این داده‌ها وجود ندارد. روش‌های انتخاب ویژگی نظارتی، هنگام مواجهه با این نوع مسائل به خوبی عمل نمی‌کنند. رویکرد جدید در مواجهه با این مسائل، انتخاب ویژگی نیمه‌نظارتی است که از داده‌های برچسب‌دار و بدون برچسب برای انتخاب ویژگی‌ها استفاده می‌کند [۱۲]. در روش‌های انتخاب ویژگی نیمه‌نظارتی، از برچسب داده‌های برچسب‌دار و اطلاعات توزیع و ساختار محلی داده‌های برچسب‌دار و بدون برچسب برای انتخاب ویژگی‌ها استفاده می‌شود [۱۵].

روش‌های انتخاب ویژگی نظارتی و نیمه‌نظارتی کلاسیک نظیر امتیاز فیشر، امتیاز لاپلاسین (Laplacian) [۱۶] و امتیاز لاپلاسین نیمه‌نظارتی [۱۷، ۱۸]، اهمیت ویژگی‌ها را به صورت جداگانه ارزیابی می‌کنند و همبستگی (Correlation) بین ویژگی‌ها را در فرآیند انتخاب ویژگی نادیده می‌گیرند. در راستای حل این مسئله، روش‌های انتخاب ویژگی تُنک (Sparse) ارائه شده‌اند تا همبستگی بین ویژگی‌ها را در هنگام انتخاب ویژگی در نظر بگیرند [۱۹–۲۱].

هدف روش‌های انتخاب ویژگی تُنک، استفاده از مدل‌های تُنک برای محاسبه ماتریس انتقال (Transformation matrix) بهینه جهت انتخاب ویژگی و دستیابی به نمایش تُنکی از داده‌ها است. با در نظر گرفتن مسئله انتخاب ویژگی، ماتریس انتقال باید تُنک باشد و تنها تعداد کمی از سطرهای آن غیر صفر باشد. در واقع، سطرهای ماتریس انتقال برای وزن‌دهی به ویژگی‌ها مورد استفاده قرار می‌گیرند و ویژگی‌های متناظر با سطرهای غیر صفر این ماتریس انتخاب می‌شوند. برای این‌که ماتریس انتقال برای مسئله انتخاب ویژگی مناسب باشد، یک عبارت منظم‌سازی (Regularization) مبتنی بر مدل‌های تُنک به تابع هدف محاسبه ماتریس انتقال اضافه می‌گردد تا سطرهای این ماتریس تُنک شوند [۱۹–۲۱].

مشهورترین مدل تُنک نُرم-l_1 (lasso) [۲۲] است و نُرم- l_p $(0 < p < 1)$ به عنوان توسعه‌ای از آن برای به دست آوردن نمایش تُنک‌تری از داده‌ها است [۲۳، ۲۴]. محققان زیادی نشان داده‌اند که روش‌های انتخاب ویژگی تُنک مبتنی بر نُرم- $l_{2,1}$ به صورت مناسب‌تری همبستگی میان ویژگی‌ها را در هنگام انتخاب ویژگی درنظر می‌گیرند [۱۹، ۲۱، ۲۵]. وانگ و همکاران، با معرفی نُرم-$l_{2,p}$ $(0 < p \leq 1)$ به عنوان توسعه‌ای از نُرم-$l_{2,1}$، آن را در یک چارچوب انتخاب ویژگی نظارتی به کار بردند و نشان دادند که نُرم-$l_{2,1/2}$ کارایی بهتری نسبت به نُرم-$l_{2,1}$ دارد و ویژگی‌ها را به صورت تُنک‌تر انتخاب می‌کند [۲۶]. اکثر روش‌های انتخاب ویژگی تُنک به صورت نظارتی عمل می‌کنند.

روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر مدل‌های تُنک، با استفاده از عبارت منظم‌سازی تُنک و اطلاعات برچسب داده‌های برچسب‌دار و اطلاعات توزیع هندسی و ساختار محلی داده‌های برچسب‌دار و بدون برچسب، ویژگی‌های مناسب را انتخاب می‌نمایند [۲۰، ۲۱]. اکثر روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر مدل‌های تُنک، از منظم‌سازی منیفلد استفاده می‌کنند تا بتوانند با در نظر گرفتن ساختار هندسی داده‌های برچسب‌دار و بدون برچسب، مناسب‌ترین ویژگی‌ها را انتخاب نمایند. منظم‌سازی مبتنی بر لاپلاسین، متداول‌ترین نمایش منیفلد داده‌ها است که از لاپلاسین گراف برای نمایش ساختار هندسی داده‌ها استفاده می‌کند. منظم‌سازی لاپلاسین توانایی برون‌یابی ضعیفی دارد و نمی‌تواند ساختار محلی داده‌ها را به خوبی حفظ کند [۲۷]. منظم‌سازی هسین در مقایسه با منظم‌سازی لاپلاسین توانایی برون‌یابی بهتری دارد و می‌تواند ساختار هندسی محلی داده‌ها را به خوبی حفظ کند. منظم‌سازی هسین می‌تواند باعث افزایش کارایی یادگیری نیمه‌نظارتی و انتخاب ویژگی نیمه‌نظارتی شود [۲۷].

در این مقاله، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر پیشنهاد می‌شود که همبستگی میان ویژگی‌ها را در هنگام انتخاب ویژگی درنظر می‌گیرد و ماتریس انتقال بهینه برای انتخاب ویژگی را محاسبه می‌نماید. روش پیشنهادی، منظم‌سازی هسین، آنالیز تشخیصی فیشر و نُرم-$l_{2,1}$ را در یک چارچوب انتخاب ویژگی نیمه‌نظارتی فیلتر ترکیب می‌کند. برای حل روش پیشنهادی، یک الگوریتم با رویکرد تکراری استفاده می‌شود و همگرایی آن به صورت تئوری و عملی اثبات می‌شود. کارایی روش پیشنهادی، با استفاده از دو روش طبقه‌بندی در مقایسه با هشت روش انتخاب ویژگی بر روی پنج مجموعه داده ارزیابی می‌شود. نتایج آزمایش‌ها، برتری روش پیشنهادی در مقایسه با سایر روش‌های انتخاب ویژگی را نشان می‌دهد.

ادامه مقاله به صورت زیر سازمان‌دهی شده است. در بخش دوم، کارهای مرتبط مرور می‌شود و در بخش سوم، به تشریح روش پیشنهادی پرداخته و برای حل آن الگوریتمی تکراری استفاده شده و همگرایی آن ثابت می‌شود. بخش چهارم به چگونگی پیاده‌سازی و بیان نتایج بر روی پنج مجموعه داده می‌پردازد. در بخش پنجم، نتیجه‌گیری مقاله ارائه می‌شود.

۲- مروری بر کارهای مرتبط

در این بخش، مروری بر روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر معیار فیشر و مدل‌های تُنک انجام می‌شود.

۲-۱- روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر معیار فیشر

امتیاز فیشر، یک روش انتخاب ویژگی نظارتی است و به جستجوی ویژگی‌ها با بهترین توانایی تفکیک‌پذیری می‌پردازد [۲۸]. می‌توان از معیار فیشر که فرمول معادل امتیاز فیشر است نیز برای محاسبه امتیاز فیشر استفاده کرد. در این روش، ماتریس انتقال برای انتخاب ویژگی به دست می‌آید که هر سطر این ماتریس اهمیت هر ویژگی را مشخص می‌کند. هدف روش‌های انتخاب ویژگی نظارتی مبتنی بر معیار فیشر بیشینه نمودن فاصله بین-کلاسی و کمینه نمودن فاصله نمونه‌های درون-کلاسی است [۲۹]. این روش‌ها برای اندازه‌گیری اهمیت هر ویژگی، ساختار هندسی داده‌ها را درنظر نمی‌گیرند.

روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر معیار فیشر، از خصوصیات معیار فیشر و اطلاعات توزیع داده‌ها و ساختار محلی آنها استفاده می‌کنند [۳۰–۳۳]. در این روش‌ها، می‌توان از امتیاز فیشر یا معیار فیشر استفاده نمود. در روش‌های نیمه‌نظارتی که از امتیاز فیشر استفاده می‌نمایند، معمولاً یک عبارت جدید بر اساس فرض منیفلد که ساختار هندسی داده‌ها را درنظر می‌گیرد به امتیاز فیشر اضافه می‌شود [۳۰، ۳۱]. در روش‌های نیمه‌نظارتی که از معیار فیشر استفاده می‌نمایند، می‌توان یک عبارت جدید براساس فرض منیفلد اضافه کرد [۳۳] یا به کمک داده‌های برچسب‌دار و بدون برچسب، ماتریس‌های پراکندگی درون-کلاسی و بین-کلاسی را تعریف نمود. برای این‌کار می‌توان گراف‌های درون-کلاسی و بین-کلاسی را ایجاد نمود و با استفاده از آنها لاپلاسین گراف‌ها و ماتریس‌های پراکندگی را به دست آورد [۳۲]. روش دیگری که برای تعیین این ماتریس‌ها به‌کار می‌رود، استفاده از روش‌های انتشار برچسب (Label propagation) برای تعیین برچسب داده‌های بدون برچسب و تعیین ماتریس‌های پراکندگی از روی برچسب‌های تعیین شده می‌باشد [۳۴،۳۵]. یانگ و همکاران با ترکیب معیار حفظ ساختار محلی و استراتژی واریانس در چارچوب امتیاز فیشر، یک روش انتخاب ویژگی نیمه‌نظارتی مبتنی بر امتیاز فیشر ارائه دادند که از دانش پیشینی مبتنی بر لاپلاسین گراف استفاده می‌کند [۳۳].

روش‌های ارائه شده فوق مبتنی بر امتیاز یا معیار فیشر، همبستگی بین ویژگی‌ها را در هنگام انتخاب ویژگی در نظر نمی‌گیرد و آنها را یکی یکی انتخاب می‌کنند.

۲-۲- روش‌های انتخاب ویژگی نیمه‌نظارتی تُنک

در سال‌های اخیر، استفاده از انواع مدل‌های تُنک برای انتخاب ویژگی‌ها و دستیابی به نمایش تنکی از داده‌ها به سرعت توسعه پیدا کرده است. هدف این روش‌ها، محاسبهٔ ماتریس انتقال $W \in R^{d\times c}$ بهینه برای انتخاب ویژگی است. تاکنون، مدل‌های تُنک نُرم-l_1 (lasso) و نُرم- l_p $(0 < p < 1)$ برای

انتخاب ویژگی و به دست آوردن نمایش تُنکی از داده‌ها مورد استفاده قرار گرفته‌اند ولی این روش‌ها، اطلاعات مفید همبستگی بین ویژگی‌ها را نادیده می‌گیرند [۲۰].

در روش‌های انتخاب ویژگی نیمه‌نظارتی، معمولاً از مدل‌های تُنک مبتنی بر نُرم- $l_{2,1}$ [۱۹، ۲۱]، نُرم-$l_{2,p}$ $(0 < p < 1)$ [۲۰، ۲۷] و نُرم-$l_{2,p}$ $(0 < p \leq 1)$ [۳۶، ۳۷] برای محاسبهٔ ماتریس انتقال بهینه تُنک استفاده می‌شود که همبستگی بین ویژگی‌ها را برای انتخاب ویژگی در نظر می‌گیرند. ما و همکاران، یک روش انتخاب ویژگی ساختاری تُنک ارائه دادند که از ترکیب یادگیری نیمه‌نظارتی، تابع زیان مبتنی بر نُرم- مرتبه دوم و عبارت منظم‌سازی مبتنی بر نُرم- $l_{2,1}$ برای انتخاب ویژگی استفاده می‌کند [۱۹]. یک روش انتخاب ویژگی تُنک مبتنی بر لاپلاسین گراف توسط شای و همکاران ارائه شد که به صورت نیمه‌نظارتی عمل کرده و از تابع زیان مبتنی بر نُرم- مرتبه دوم و عبارت منظم‌سازی مبتنی بر نُرم-$l_{2,1/2}$ برای انتخاب ویژگی استفاده می‌کند [۲۰]. شیخ‌پور و همکاران، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک برای مسائل رگرسیون ارائه دادند که از تابع هدفی مبتنی بر نُرم-l_{2p} $(0 < p \leq 1)$ و ماتریس پراکندگی نیمه‌نظارتی مبتنی بر لاپلاسین گراف استفاده می‌کند [۳۷]. یک روش انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر لاپلاسین گراف توسط شیخ‌پور و همکاران ارائه شد که از نُرم-$l_{2,p}$ $(0 < p \leq 1)$ بر روی تابع زیان و عبارت منظم‌سازی استفاده می‌کند [۳۶]. لی و همکاران، یک روش انتخاب ویژگی نیمه‌نظارتی مبتنی بر محدودیت عمومی غیروابسته ارائه دادند که از لاپلاسین گراف و عبارت منظم‌سازی مبتنی بر نُرم- $l_{2,1}$ برای انتخاب ویژگی‌های متمایز و غیروابسته استفاده می‌کند [۳۸].

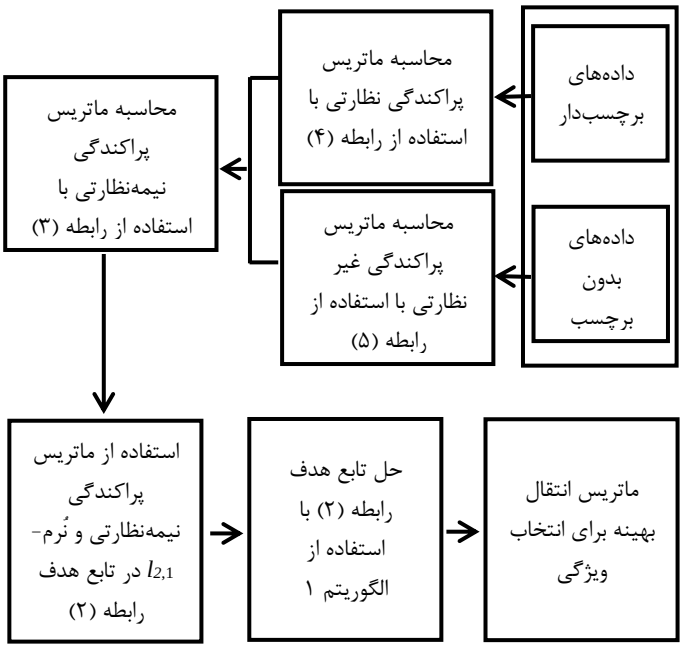
روش‌های انتخاب ویژگی نیمه‌نظارتی تُنک ذکر شده، از منظم‌سازی لاپلاسین که متداول‌ترین روش برای در نظر گرفتن ساختار هندسی داده‌های برچسب‌دار و بدون برچسب و نمایش منیفلد داده‌ها است، استفاده می‌نمایند. به دلیل توانایی ضعیف برون‌یابی منظم‌سازی لاپلاسین و عدم حفظ مناسب ساختار محلی داده‌ها توسط آن، روش‌های انتخاب ویژگی نیمه‌نظارتی مبتنی بر لاپلاسین گراف ممکن است نتوانند هنگامی که داده‌های برچسب‌دار کمی در دسترس هستند، کارایی خوبی داشته باشند [۲۷].

منظم‌سازی هسین، با توانایی برون‌یابی بهتر نسبت به منظم‌سازی لاپلاسین و حفظ مناسب ساختار هندسی داده‌ها، می‌تواند باعث افزایش کارایی انتخاب ویژگی نیمه‌نظارتی شود [۲۷]. شای و همکاران، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک ارائه دادند که یادگیری نیمه‌نظارتی مبتنی بر منظم‌سازی هسین، تابع زیان مبتنی بر نُرم- مرتبه دوم و عبارت منظم‌سازی مبتنی بر نُرم-$l_{2,1/2}$ را در یک تابع هدف ترکیب می‌کند [۲۷].

هان و همکاران، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر نُرم- $l_{2,1}$ ارائه دادند که از ترکیب ماتریس پراکندگی نظارتی مبتنی بر معیار فیشر و ماتریس پراکندگی نیمه‌نظارتی مبتنی بر معیار اسپیلاین برای حفظ ساختار محلی داده‌ها استفاده می‌کند [۲۱].

۳- روش پیشنهادی

در این بخش، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر ماتریس هسین و آنالیز تشخیصی فیشر[1] (SFSHF) برای محاسبه ماتریس انتقال بهینه پیشنهاد می‌شود که از نُرم-$l_{2,1}$ برای انتخاب ویژگی تُنک استفاده می‌کند. در روش پیشنهادی، تابع هدفی مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر ارائه می‌شود که با استفاده از اطلاعات تفکیکی داده‌های برچسب‌دار و ساختار محلی داده‌های برچسب‌دار و بدون برچسب و با در نظر گرفتن همبستگی میان ویژگی‌ها، انتخاب ویژگی‌های مناسب را انجام می‌دهد. استفاده از منظم‌سازی هسین در روش پیشنهادی باعث می‌شود ساختار هندسی محلی داده‌ها به خوبی حفظ شود و استفاده از نُرم-$l_{2,1}$ باعث می‌شود همبستگی میان ویژگی‌ها در نظر گرفته شود. فلوچارت روش پیشنهادی در شکل ۱ نشان داده شده است.



شکل ۱- فلوچارت روش پیشنهادی برای محاسبه ماتریس انتقال بهینه برای انتخاب ویژگی

قبل از بررسی روش پیشنهادی، نمادهایی که در این بخش برای تعریف رابطه‌ها به کار رفته‌اند بیان می‌شوند. مجموعه‌ای از n نمونه آموزشی $X = [x_1, \dots x_l, , x_{l+1}, \dots, x_n] \in R^{d\times n}$ شامل l نمونه برچسب‌دار و $(n\text{-}l)$ نمونه بدون برچسب را در نظر بگیرید که هر نمونه با d ویژگی مشخص می‌شود و $x_i \in R^d$ $(1 \leq i \leq n)$، iامین نمونه آموزشی را نشان می‌دهد. مجموعه $Y_l = [y_1, \dots, y_l]^T \in \{0,1\}^{l\times c}$ برچسب داده‌ها را مشخص می‌کند که c تعداد کلاس‌ها را بیان می‌کند و $y_i \in R^c$ $(1 \leq i \leq l)$، بردار برچسب iامین نمونه آموزشی را مشخص می‌کند. اگر نمونه x_i متعلق به jامین کلاس باشد آنگاه $y_{ij} = 1$، در غیر این صورت $y_{ij} = 0$.

فرض کنید $W \in R^{d\times c}$ ماتریس انتقالی باشد که برای انتخاب ویژگی تُنک استفاده می‌شود. نُرم -$l_{2,1}$ این ماتریس به صورت زیر تعریف می‌شود:

$$\|W\|_{2,1} = \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} W_{i,j}^2} = \sum_{i=1}^{d} \|w^i\|_2 \qquad (۱)$$

در چارچوب کلی کاهش ابعاد، ماتریس انتقال می‌تواند با استفاده از کمینه‌سازی $Tr\ (W^T MW)$ به دست آید که ماتریس M می‌تواند به گونه‌ای تعریف شود که ساختارهای خاصی از داده‌های آموزشی را بیان کند. با ترکیب انتخاب ویژگی مبتنی بر نُرم-$l_{2,1}$ با کمینه‌سازی $Tr\ (W^T\ MW)$، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک برای به دست آوردن ماتریس انتقال بهینه برای انتخاب ویژگی نیمه‌نظارتی می‌تواند به صورت حل مسئله زیر باشد [۲۱، ۳۹]:

$$\min_{W^TW=1} Tr(W^T MW) + \lambda\|W\|_{2,1} \qquad (۲)$$

که عبارت منظم‌سازی $\|W\|_{2,1}$ باعث می‌شود که سطرهای ماتریس W

[1] Semisupervised Feature Selection based on Hessian regularization and Fisher discriminant analysis (SFSHF)

تُنک شود و برای انتخاب ویژگی مناسب باشد. همچنین این عبارت باعث می شود که همبستگی میان ویژگی‌ها در هنگام انتخاب ویژگی در نظر گرفته شود. در رابطه فوق، پارامتر $\lambda > 0$ اثر منظم‌سازی را کنترل می‌کند و $M \in R^{d\times d}$ یک ماتریس پراکندگی (Scatter matrix) نیمه‌نظارتی است که توزیع داده‌ها و اطلاعات برچسب را بیان می‌کند. در این مقاله، یک ماتریس پراکندگی نیمه‌نظارتی مبتنی بر آنالیز تشخیصی فیشر و ماتریس هسین پیشنهاد می‌شود که اطلاعات برچسب و ساختار هندسی محلی داده‌ها را بیان می‌کند. ماتریس پراکندگی نیمه‌نظارتی M به صورت زیر تعریف می‌شود:

$$M = S + \mu U \tag{۳}$$

پارامتر μ $(0 \le \mu \le 1)$، وزن ماتریس U را کنترل می‌کند.

ماتریس $S \in R^{d\times d}$ یک ماتریس پراکندگی نظارتی است که از داده‌های برچسب‌دار به دست می‌آید. ماتریس $U \in R^{d\times d}$ یک ماتریس پراکندگی غیرنظارتی است که اطلاعات ساختار محلی همه داده‌های آموزشی را بیان می‌کند. ماتریس پراکندگی نیمه‌نظارتی پیشنهادی در این مقاله، ترکیب دو ماتریس پراکندگی نظارتی مبتنی بر آنالیز تشخیصی فیشر و ماتریس پراکندگی غیرنظارتی مبتنی بر منظم‌سازی هسین است. هنگامی‌که W بهینه به دست آید، ویژگی‌ها با توجه به مقدار $\|w_i\|_2$ $(i = 1,..,d)$ به ترتیب نزولی مرتب شده و ویژگی‌ها با رتبه بالاتر انتخاب می‌شوند. در ادامه، ماتریس‌های پراکندگی نظارتی و غیرنظارتی به کار رفته در این مقاله شرح داده می‌شوند.

۳-۱- ماتریس پراکندگی نظارتی مبتنی بر آنالیز تشخیصی فیشر

آنالیز تشخیصی فیشر، یکی از روش‌های متداول برای استفاده از اطلاعات جداپذیری داده‌های برچسب‌دار برای جستجوی ویژگی‌هایی است که بهترین توانایی تفکیک‌پذیری را دارا می‌باشند [۲۸]. هدف روش‌های مبتنی بر آنالیز تشخیصی فیشر بیشینه نمودن نسبت ماتریس‌های پراکندگی بین-کلاسی و درون-کلاسی است. در این روش‌ها، فاصله نمونه‌های درون کلاس‌ها کمینه شده و فاصله نمونه‌های بین کلاس‌های مختلف بیشینه می‌شود.

در این مقاله، از ماتریس پراکندگی درون کلاسی برای استفاده از اطلاعات برچسب داده‌ها به صورت زیر استفاده می‌شود:

$$\mathrm{S} = \sum_{i=1}^{c} \frac{1}{n_i} \left(\sum_{j=1}^{n_i} (x_j^i - \mu^i)(x_j^i - \mu^i)^T \right) \tag{۴}$$

در رابطه فوق، μ^i بردار میانگین کلاس iام، n_i تعداد نمونه‌های برچسب‌دار در کلاس iام، و x_j^i نمونه jام در کلاس iام را مشخص می‌کند.

۳-۲- ماتریس پراکندگی غیرنظارتی مبتنی بر منظم‌سازی هسین

ماتریس پراکندگی غیرنظارتی U، با استفاده از همه داده‌های آموزشی به صورت زیر محاسبه می‌شود.

$$U = XHX^T \tag{۵}$$

که $H \in \mathrm{R}^{n\times n}$ ماتریس به دست آمده از منظم‌سازی هسین است که با استفاده از همه داده‌های آموزشی به دست می‌آید و ساختار هندسی محلی داده‌ها را بیان می‌کند. برای محاسبه ماتریس هسین، برچسب داده‌ها نادیده گرفته می‌شود.

در ادامه، منظم‌سازی هسین معرفی شده در [۴۰] شرح داده می‌شود. یک منیفلد هموار $F \subset R^n$ و یک فضای مماس محلی T_x^*F برای نقطه $x \in F$ در نظر بگیرید. همچنین، فرض کنید $\varphi(x_i) = f_i$ تابع نگاشتی است که برچسب f_i مرتبط با داده x_i را پیش‌بینی می‌کند. بنابراین، انرژی هسین $E_{H(\varphi)}$ برای تابع $\varphi: F \to R$ به صورت زیر تعریف می‌شود:

$$\mathrm{E}_H(\varphi) = \int_F \|\nabla_a \nabla_b \varphi\|^2_{T_x^*F \otimes T_x^*F} \, dV(x) \tag{۶}$$

در رابطه فوق، $\nabla_a \nabla_b \varphi$ مشتق کوواریانت مرتبه دوم φ و $dV(x)$ عنصری روی منیفلد F است.

در مختصات نرمال، رابطه زیر برقرار است:

$$\|\nabla_a \nabla_b \varphi\|^2_{T_x^*F \otimes T_x^*F} = \sum_{r,s=1}^{l} \left(\frac{\partial^2 \varphi}{\partial x^r \partial x^s} \right)^2 \tag{۷}$$

به طوری که در یک نقطه معین x، نُرم مشتق کواریانت مرتبه دوم معادل نُرم فروبنیوس (Frobenius norm) هسین φ در مختصات نرمال است.

فرض کنید $N_k(x_i)$، مجموعه‌ای از k-نزدیکترین همسایه x_i است. با توجه به مقادیر تابع $\varphi(x_t) = f_t$ بر روی $N_k(x_i)$، هسین φ در x_i می‌تواند به صورت رابطه زیر نوشته شود:

$$\left. \frac{\partial^2 \varphi}{\partial x^r \partial x^s} \right|_{x_i \approx \sum_{t=1}^{k} Z_{rst}^{(i)} f_t} \tag{۸}$$

در رابطه فوق، $Z_{rst}^{(i)}$ عملگر هسین است که می‌تواند با استفاده از چند جمله‌ای مرتبه دوم با استفاده از حداقل مربعات خطی محاسبه شود.

فرض کنید $\mathrm{f}_\alpha = \varphi(x_\alpha)$ و $\mathbf{f} = (\mathrm{f}_1, \dots, \mathrm{f}_k)^T \in R^k$. بنابراین، تخمین نُرم فروبنیوس هسین φ در x_i با استفاده از رابطه زیر به دست می‌آید:

$$\|\nabla_a \nabla_b \varphi\|^2 \approx \sum_{r,s=1}^{l} \left(\sum_{\alpha=1}^{k} Z_{rs\alpha}^{(i)} \mathrm{f}_\alpha \right)^2 = \sum_{\alpha,\beta=1}^{k} \mathrm{f}_\alpha \mathrm{f}_\beta H_{\alpha\beta}^{(i)} \tag{۹}$$

که $H_{\alpha\beta}^{(i)} = \sum_{r,s=1}^{l} Z_{rs\alpha}^{(i)} Z_{rs\beta}^{(i)}$، بنابرای انرژی هسین می‌تواند به صورت رابطه زیر نوشته می‌شود.

$$\hat{E}_{H(\varphi)} = \sum_{i=1}^{n} \sum_{\alpha \in N_k(x_i)} \sum_{\beta \in N_k(x_i)} \mathrm{f}_\alpha \mathrm{f}_\beta H_{\alpha\beta}^{(i)} = F^T H F \tag{۱۰}$$

که $H = \sum_{i=1}^{n} H^{(i)}$ و n تعداد داده‌های آموزشی است.

منظم‌سازی هسین در مقایسه با منظم‌سازی لاپلاسین ساختار منیفلد محلی داده‌ها را بهتر حفظ می‌کند [۲۷]. بنابراین، منظم‌سازی هسین می‌تواند برای بالابردن کارایی یادگیری نیمه‌نظارتی و انتخاب ویژگی نیمه‌نظارتی به کار رود.

۳-۳- حل روش پیشنهادی SFSHF

در این بخش، برای حل تابع هدف رابطه (۲) با وجود قید تساوی $W^TW = I$، الگوریتمی استفاده می‌شود که بتواند این مسئله کمینه‌سازی را حل نماید. نقاط بهینه مسائل بهینه‌سازی با وجود قیود تساوی می‌توانند با استفاده از روش ضرایب لاگرانژ به دست آیند.

فرض کنید $W = [w_1, \dots, w_d]^T$، بنابراین $\|W\|_{2,1} = Tr(W^T D W)$، که D یک ماتریس قطری است که عناصر قطر اصلی آن به صورت $D^{ii} = \frac{1}{2\|w_i\|_2}$ محاسبه می‌شوند.

تابع هدف رابطه (۲) معادل با رابطه زیر است:

$$\min_{W^TW=I} Tr(W^T M W) + \lambda Tr(W^T D W) \tag{۱۱}$$

با توجه به مسئله لاگرانژ، تابع هدف رابطه (۱۱) به صورت زیر تبدیل می شود:

$$\min_{W} Tr(W^T M W) + \lambda Tr(W^T D W) - Tr\big(\gamma(W^T W - I)\big) \tag{۱۲}$$

از تابع فوق نسبت به W مشتق گرفته و آن را مساوی صفر قرار می‌دهیم.

$$2MW + 2\lambda DW - 2\gamma W = 0, \quad W(M + \lambda D) = \gamma W \tag{۱۳}$$

رابطه فوق یک مسئله ویژه (Eigen problem) است که برای به دست

آوردن ماتریس W برای انتخاب ویژگی، باید تجزیه ویژه[۲] $M+\lambda D$ را انجام دهیم.

الگوریتم ۱ برای حل تابع هدف رابطه (۲) به کار می‌رود.

الگوریتم ۱: الگوریتم استفاده شده برای حل روش پیشنهادی SFSHF

ورودی:

ماتریس مجموعه داده‌های آموزشی برچسب‌دار $X_l \in R^{d\times l}$ که l تعداد نمونه‌های برچسب‌دار و d ابعاد هر نمونه می‌باشد.

ماتریس مجموعه داده‌های آموزشی برچسب‌دار و بدون برچسب $X \in R^{d\times n}$ که n تعداد نمونه‌ها و d ابعاد هر نمونه است.

ماتریس برچسب داده‌های آموزشی برچسب‌دار $Y_l \in R^{l\times c}$.

پارامتر منظم‌سازی λ

خروجی:

ایندکس c ویژگی انتخاب شده با رتبه بالاتر

بردار $W \in R^{d\times c}$ همگرا شده

روش:

ماتریس پراکندگی نظارتی مبتنی بر آنالیز تشخیصی فیشر $\mathbf{S}$ با استفاده از رابطه (۴) محاسبه شود.

ماتریس پراکندگی غیرنظارتی مبتنی بر منظم‌سازی هسین $\boldsymbol{U}$ با استفاده از رابطه (۵) محاسبه شود.

ماتریس پراکندگی نیمه‌نظارتی M با استفاده از رابطه (۳) محاسبه شود.

مقدار t مساوی صفر قرار داده شود (t = 0) و بردار $\boldsymbol{W^t} \in \boldsymbol{R^d}$ به صورت تصادفی مقداردهی اولیه شود.

مراحل زیر تکرار شوند:

ماتریس قطری $\boldsymbol{D^t} \in \boldsymbol{R^{d\times d}}$ به صورت $\boldsymbol{D^{ii}} = \frac{\mathbf{1}}{\mathbf{2}\|\boldsymbol{w_i^t}\|_2}$ محاسبه شود.

ماتریس $\boldsymbol{R^t} \in \boldsymbol{R^{d\times d}}$ به صورت $\boldsymbol{R^t} = \boldsymbol{M} + \boldsymbol{\lambda D^t}$ محاسبه شود.

ماتریس $\boldsymbol{W^{t+1}}$ مطابق $\boldsymbol{W^{t+1}} = [\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_q}]$ به روزرسانی شود، به طوری که $\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_q}$ بردارهای ویژه متناظر با q تا ازکوچکترین مقادیر ویژه ماتریس $\boldsymbol{R^t}$ هستند.

t=t+1

تا زمانی‌که مقدار تابع هدف همگرا شود.

ویژگی‌ها با توجه به $\|w_i\|_2$ به ترتیب نزولی مرتب شده و c ویژگی با رتبه بالاتر انتخاب شوند.

همان‌گونه که در شکل ۱ و الگوریتم ۱ بیان شد، داده‌های آموزشی به دو دسته برچسب‌دار و بدون برچسب تقسیم می‌شوند و ماتریس پراکندگی نظارتی با استفاده از داده‌های برچسب‌دار و ماتریس پراکندگی غیرنظارتی با استفاده از داده‌های بدون برچسب به ترتیب مطابق رابطه‌های (۴) و (۵) محاسبه می‌شوند. سپس، ماتریس پراکندگی نیمه‌نظارتی مبتنی بر ماتریس‌های پراکندگی نظارتی و غیرنظارتی با استفاده از رابطه (۳) محاسبه می‌شود. پس از به دست آمدن ماتریس انتقال بهینه W با استفاده از رویکرد تکراری الگوریتم ۱، هریک از سطرهای این ماتریس برای وزن‌دهی به ویژگی‌ها مورد استفاده قرار می‌گیرند و ویژگی‌ها با رتبه بالاتر انتخاب می‌شوند.

[۲] Eigen-decomposition

۳-۴- آنالیز همگرایی

در این بخش مشابه آنالیز همگرایی الگوریتم ارائه شده در [۲۱]، با استفاده از تئوری ۱ اثبات می‌شود که رویکرد تکراری در الگوریتم ۱ به W بهینه همگرا می‌شود. قبل از بیان تئوری۱، ابتدا دو لم ارائه شده در [۲۶] معرفی می‌شوند:

لم ۱: اگر $\varphi(t) = 2t - t^2 - 1$، آنگاه برای هر مقدار $t > 0$، $\varphi(t) \le 0$.

لم ۲: فرض کنید که بردارهای g_i^t و g_i^{t+1}، به‌ترتیب iامین سطر ماتریس‌های G^t و G^{t+1} هستند. بنابراین نامساوی زیر برقرار است.

$$\|g_i^{t+1}\|_2 - \frac{\|g_i^{t+1}\|_2^2}{2\|g_i^t\|_2} \le \|g_i^t\|_2 - \frac{\|g_i^t\|_2^2}{2\|g_i^t\|_2}\ ,\ i = 1, \dots, r$$

می‌توان رابطه زیر را از نامساوی فوق نتیجه گرفت.

$$\sum_{i=1}^{r}\left(\|g_i^{t+1}\|_2 - \frac{\|g_i^{t+1}\|_2^2}{2\|g_i^t\|_2}\right) \le \sum_{i=1}^{r}\left(\|g_i^t\|_2 - \frac{\|g_i^t\|_2^2}{\|g_i^t\|_2}\right)$$

تئوری ۱: مقدار تابع هدف در رابطه (۲) به طور یکنواخت در هر گام کاهش می‌یابد تا زمانی‌که همگرا شود.

اثبات: با توجه به الگوریتم ۱، می‌توان از رابطه (۱۱) نتیجه گرفت که:

$$W^{t+1} = \min_{W^TW=1} Tr(W^TMW) + \lambda Tr(W^TDW)$$

از آنجایی‌که W^{t+1} کمینه رابطه $Tr(W^TMW) + \lambda Tr(W^TDW)$ است، بنابراین:

$$Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda Tr\left(W^{t+1^T}D^tW^{t+1}\right) \le Tr\left(W^{t^T}D^tW^t\right) + \lambda Tr\left(W^{t^T}D^tW^t\right)$$

$$Tr\left(w^{t^T}D^tw^t\right) = \sum_{i=1}^{d}\frac{\|w_i^t\|_2^2}{2\|w_i^t\|_2}\quad ,\ Tr\left(w^{t+1^T}D^tw^{t+1}\right) = \sum_{i=1}^{d}\frac{\|w_i^{t+1}\|_2^2}{2\|w_i^t\|_2}$$

$$\Rightarrow Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda\sum_{i=1}^{d}\frac{\|w_i^{t+1}\|_2^2}{2\|w_i^t\|_2} \le Tr\left(W^{t^T}D^tW^t\right) + \lambda\sum_{i=1}^{d}\frac{\|w_i^t\|_2^2}{2\|w_i^t\|_2}$$

$$\Rightarrow Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda\sum_{i=1}^{d}\frac{\|w_i^{t+1}\|_2^2}{2\|w_i^t\|_2} - \lambda\sum_{i=1}^{d}\|w_i^{t+1}\|_2 + \lambda\sum_{i=1}^{d}\|w_i^{t+1}\|_2 \le Tr\left(W^{t^T}D^tW^t\right) + \lambda\sum_{i=1}^{d}\frac{\|w_i^t\|_2^2}{2\|w_i^t\|_2} - \lambda\sum_{i=1}^{d}\|w_i^t\|_2 + \lambda\sum_{i=1}^{d}\|w_i^t\|_2$$

$$\Rightarrow Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda\sum_{i=1}^{d}\|w_i^{t+1}\|_2 - \lambda\left(\sum_{i=1}^{d}\|w_i^{t+1}\|_2 - \sum_{i=1}^{d}\frac{\|w_i^{t+1}\|_2^2}{2\|w_i^t\|_2}\right) \le Tr\left(W^{t^T}D^tW^t\right) + \lambda\sum_{i=1}^{d}\|w_i^t\|_2 - \lambda\left(\sum_{i=1}^{d}\|w_i^t\|_2 - \sum_{i=1}^{d}\frac{\|w_i^t\|_2^2}{2\|w_i^t\|_2}\right)$$

با توجه به لم ۲

$$\sum_{i=1}^{r}\left(\|w_i^{t+1}\|_2 - \frac{\|w_i^{t+1}\|_2^2}{2\|w_i^t\|_2}\right) \le \sum_{i=1}^{r}\left(\|w_i^t\|_2 - \frac{\|w_i^t\|_2^2}{2\|w_i^t\|_2}\right)$$

بنابراین

$$Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda\sum_{i=1}^{d}\|w_i^{t+1}\|_2 \le Tr\left(W^{t^T}D^tW^t\right) + \lambda\sum_{i=1}^{d}2\|w_i^t\|_2$$

$$\Rightarrow Tr\left(W^{t+1^T}MW^{t+1}\right) + \lambda\|w_i^{t+1}\|_2 \le Tr\left(W^{t^T}D^tW^t\right) + \lambda\|w_i^t\|_{2,1}$$

که مشخص می‌کند مقدار تابع هدف در رابطه (۲) به طور یکنواخت در هر گام کاهش می‌یابد تا همگرا شود.

۴- شبیه‌سازی

برای ارزیابی کارایی روش انتخاب ویژگی پیشنهادی، آزمایش‌هایی انجام می‌شود و کارایی آن با تعدادی روش انتخاب ویژگی مقایسه می‌گردد.

۴-۱- روش‌های مورد مقایسه

در این بخش، کارایی روش پیشنهادی با چندین روش انتخاب ویژگی

تُنک مقایسه می‌شود. این روش‌ها عبارتند از:

انتخاب ویژگی تُنک غیرنظارتی مبتنی بر نُرم-$l_{2,1}$ (EGCFS) [۴۱]: روش EGCFS، یک روش انتخاب ویژگی غیرنظارتی تُنک است که مبتنی بر یادگیری گراف تطبیقی و نُرم-$l_{2,1}$ است. این روش، ایده بیشینه نمودن ماتریس پراکندگی بین-کلاسی و ساختار گراف تطبیقی را در یک تابع هدف ترکیب می‌کند تا ویژگی‌های غیروابسته‌ی متمایز را انتخاب نماید.

انتخاب ویژگی جداکننده پراکندگی زیرفضا (S^{2}DFS)[۳] [۴۲]: روش S^{2}DFS، یک روش انتخاب ویژگی نظارتی مبتنی بر نُرم-$l_{2,0}$ است که از تابع هدفی مبتنی بر آنالیز جداکننده خطی و محدودیت زیرفضای تُنک ساختاری برای انتخاب ویژگی‌های مفید استفاده می‌کند. در این روش، با به کار بردن ایده کمینه نمودن فاصله نمونه‌های درون-کلاسی، بیشینه نمودن فاصله نمونه‌های بین-کلاسی و محدودیت زیرفضای تُنک ساختاری در یک چارچوب، ویژگی‌های مناسب انتخاب می‌شوند.

انتخاب ویژگی تُنک مبتنی بر کمینه‌سازی نُرم-$l_{2,1/2}$ (SFSN)[۴] [۲۶]: روش SFSN، یک روش انتخاب ویژگی تُنک و پایدار است که کمینه‌سازی نُرم-$l_{2,1/2}$ را بر روی تابع زیان و عبارت منظم‌سازی به کار می‌گیرد. این روش، برای انتخاب ویژگی‌ها به صورت نظارتی عمل کرده و همبستگی بین آنها را در نظر می‌گیرد.

انتخاب ویژگی با استفاده از کمینه‌سازی نُرم-$l_{2,1}$ (FSNM)[۵] [۲۵]: روش FSNM، یک روش انتخاب ویژگی تُنک مبتنی بر نُرم-$l_{2,1}$ است که به‌صورت نظارتی عمل می‌کند. این روش، کمینه‌سازی نُرم- $l_{2,1}$ را بر روی تابع زیان و عبارت منظم‌سازی اعمال می‌کند. روش FSNM، همبستگی بین ویژگی‌ها را در هنگام انتخاب ویژگی در نظر می‌گیرد.

انتخاب ویژگی تُنک مبتنی بر ماتریس اسپلاین (S^{2}FS^{2}R)[۶] [۲۱]: روش S^{2}FS^{2}R، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک است که با استفاده از نُرم-$l_{2,1}$ همبستگی بین ویژگی‌ها را در هنگام انتخاب ویژگی در نظر می‌گیرد. در این روش، از تابع هدفی همانند رابطه (۲) استفاده می‌شود و ماتریس پراکندگی نیمه‌نظارتی با استفاده از ماتریس اسپلاین و آنالیز تشخیصی فیشر به دست می‌آید.

انتخاب ویژگی تُنک هسین مبتنی بر نُرم- $l_{2,1/2}$ (HFSL)[۷] [۲۷]: این روش انتخاب ویژگی، به صورت نیمه‌نظارتی عمل کرده و از یادگیری نیمه‌نظارتی مبتنی بر منظم‌سازی هسین، تابع زیان مبتنی بر نُرم- مرتبه دوم و عبارت منظم‌سازی مبتنی بر نُرم-$l_{2,1/2}$ استفاده می‌کند.

انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر گراف[۸] (GS^{3}FS- $l_{2,1}$) [۳۶]: این روش، از یادگیری نیمه‌نظارتی مبتنی بر لاپلاسین گراف استفاده نموده و در یک چارچوب نیمه‌نظارتی، کمینه‌سازی نُرم- $l_{2,1}$ را بر روی تابع زیان و عبارت منظم‌سازی به کار می‌برد.

انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر نُرم- $l_{2,1}$ (SFS)[۹] [۳۸]: روشSFS، یک روش انتخاب ویژگی نیمه‌نظارتی مبتنی بر محدودیت عمومی غیروابسته است که با استفاده از لاپلاسین گراف و عبارت منظم‌سازی مبتنی بر نُرم-$l_{2,1}$ ویژگی‌های متمایز و غیروابسته را انتخاب می‌کند.

۴-۲- نتایج ارزیابی

برای ارزیابی کارایی روش پیشنهادی، آزمایش‌هایی بر روی پنج مجموعه داده از پایگاه UCI انجام می‌شود. مشخصات مجموعه داده‌های استفاده شده در جدول ۱ نشان داده شده است. برای آزمایش‌ها، مقادیر ویژگی‌ها در بازه [۰،۱] نرمالیزه می‌شوند. پس از انتخاب مناسب‌ترین ویژگی‌ها توسط روش‌های انتخاب ویژگی، با استفاده از روش‌های ماشین بردار پشتیبان (SVM) و k نزدیکترین همسایه برای طبقه‌بندی داده‌ها، دقت (accuracy) هر روش به دست می‌آید.

جدول ۱- مجموعه داده‌های استفاده شده

مجموعه داده	تعداد نمونه‌ها	تعداد نمونه‌های غیربرچسب‌دار	تعداد نمونه‌های مجموعه آموزشی	تعداد نمونه‌های مجموعه آزمون	تعداد ویژگی‌ها
دیابت	۷۶۸	۳۷۵	۱۲۵	۲۶۸	۸
WBCD	۶۸۳	۳۰۸	۱۰۲	۲۷۳	۹
WDBC	۵۶۹	۲۸۲	۹۴	۱۹۳	۳۰
QSAR	۱۰۵۵	۵۵۳	۱۸۵	۳۱۷	۴۱
پارکینسون	۱۹۵	۱۰۲	۳۴	۵۹	۲۲

برای انتخاب ویژگی نیمه‌نظارتی با استفاده از داده‌های برچسب‌دار و بدون برچسب، داده‌ها به سه دسته داده‌های آموزشی بدون برچسب، داده‌های آموزشی برچسب‌دار و داده‌های آزمون تقسیم می‌شوند. داده‌های غیربرچسب‌دار استفاده شده در [۱۲] به عنوان مجموعه داده‌های غیربرچسب‌دار برای مجموعه داده‌های دیابت، WBCD و WDBC در این مقاله در نظر گرفته می‌شوند و برای مجموعه داده‌های QSAR و پارکینسون، ۷۵ درصد داده‌ها به صورت داده‌های بدون برچسب به صورت تصادفی انتخاب می‌شوند. برای انتخاب ویژگی غیرنظارتی و نیمه‌نظارتی از هر دو مجموعه داده‌های آموزشی برچسب‌دار و بدون برچسب در فرایند انتخاب ویژگی استفاده می‌شود. مجموعه‌های آموزشی و آزمون برای هر مجموعه داده، سه بار ایجاد می‌شوند و آزمایش‌هایی بر روی هر کدام از این سه مجموعه داده به صورت جداگانه انجام می‌شوند و مقادیر میانگین و انحراف معیار گزارش می‌شوند.

در روش‌های انتخاب ویژگی نظارتی و طبقه‌بندی با استفاده از روش‌های ماشین بردار پشتیبان و k-نزدیکترین همسایه، تنها از نمونه‌های برچسب‌دار استفاده می‌شود. در آزمایش‌ها، چند پارامتر باید تعیین یا مقداردهی شوند که مطابق اکثر کارهای پیشین تعیین و مقداردهی می‌شوند. پارامتر k که تعداد همسایه‌ها را در روش طبقه‌بندی k-نزدیکترین همسایه مشخص می‌کند، به پنج مقداردهی می‌شود. پارامتر μ در رابطه (۳) به یک مقداردهی می‌شود تا اثر ماتریس‌های پراکندگی نظارتی و غیرنظارتی در محاسبه ماتریس پراکندگی نیمه‌نظارتی یکسان در نظر گرفته شود. تعداد ویژگی‌های انتخاب شده و پارامترهای دیگر در آزمایش‌ها، با استفاده از اعتبارسنجی متقاطع ۱۰-دسته ای (10-fold cross validation) بر روی مجموعه‌های آموزشی تعیین می‌شوند.

مقادیر بهینه پارامتر منظم‌سازی λ در روش پیشنهادی در رابطه (۲) و پارامترهای منظم‌سازی در دیگر روش‌های انتخاب ویژگی با استفاده از روش اعتبارسنجی متقاطع ۱۰-دسته‌ای و مجموعه {۰/۰،۰۰۱/۰،۰۱/۱۰۰۰،۱۰۰،۱۰،۱،۱} بر روی مجموعه داده‌های آموزشی

---

[۳] Subspace Sparsity Discriminant Feature Selection

[۴] Sparse Feature Selection based on $l_{2,1/2}$-Norms Minimization

[۵] Feature Selection via Joint $l_{2,1}$-Norms Minimization

[۶] Sparse Feature Selection based on Graph Laplacian

[۷] Hessian Sparse Feature Selection based on $l_{2,1/2}$-matrix norm

[۸] Graph based Semi-Supervised Sparse Feature Selection

[۹] Semi-Supervised Feature Selection

تعیین می‌شوند.

جدول ۲، نتایج دقت طبقه‌بندی روش‌های مختلف انتخاب ویژگی با استفاده از روش ماشین بردار پشتیبان را نشان می‌دهد.

جدول ۲- دقت طبقه‌بندی (میانگین± انحراف معیار) روش‌های مختلف انتخاب ویژگی با روش ماشین بردار پشتیبان

مجموعه داده	EGCFS	S^2DFS	SFSN	FSNM	S^2FS^2R	HFSL	GS^3FS- $l_{2,1}$	SFS	SFSHF
دیابت	۰/۷۳۰± ۰/۰۵۵	۰/۷۳۴± ۰/۰۶۷	۰/۷۳۳± ۰/۰۷۲	۰/۶۹۰± ۰/۰۷۸	۰/۷۳۸± ۰/۰۵۷	۰/۷۳۵± ۰/۰۶۲	۰/۷۳۸± ۰/۰۶۵	۰/۷۰۷± ۰/۰,۴۷	**۰/۷۴۳± ۰/۰۶۲**
WBCD	۰/۹۵۸± ۰/۰۰۸	۰/۹۵۷± ۰/۰۰۶	۰/۹۵۹± ۰/۰۰۸	۰/۹۴۶± ۰/۰۱۴	۰/۹۵۹± ۰/۰۱۴	۰/۹۶۰± ۰/۰۰۶	۰/۹۶۷± ۰/۰۰۰	۰/۹۶۵± ۰/۰۱۵	**۰/۹۷۳± ۰/۰۱۱**
WDBC	۰/۹۳۱± ۰/۰۱۶	۰/۹۱۹± ۰/۰۱۱	۰/۹۴۵± ۰/۰۲۴	۰/۹۴۶± ۰/۰۱۱	۰/۹۲۱± ۰/۰۱۶	۰/۹۴۰± ۰/۰۲۳	۰/۹۴۳± ۰/۰۱۴	۰/۹۱۷± ۰/۰۲۷	**۰/۹۴۸± ۰/۰۰۹**
QSAR	۰/۸۰۹± ۰/۰۲۲	۰/۸۰۰± ۰/۰۰۹	۰/۷۹۹± ۰/۰۰۵	۰/۸۱۲± ۰/۰۲۴	۰/۸۲۰± ۰/۰۱۱	۰/۸۰۷± ۰/۰۱۲	۰/۸۲۲± ۰/۰۲۲	۰/۸۱۲± ۰/۰۰۷	**۰/۸۳۵± ۰/۰۰۷**
پارکینسون	۰/۸۱۴± ۰/۰۱۷	۰/۸۴۸± ۰/۰۳۴	۰/۸۰۸± ۰/۰۲۶	۰/۸۲۵± ۰/۰۳۹	۰/۸۴۲± ۰/۰۳۵	۰/۸۳۰± ۰/۰۱۷	۰/۸۶۴± ۰/۰۱۷	۰/۸۱۴± ۰/۰۱۷	**۰/۸۷۶± ۰/۰۲۰**

همان‌گونه که در جدول ۲ نشان داده شده است، روش انتخاب ویژگی پیشنهادی SFSHF با استفاده از روش ماشین بردار پشتیبان دارای کارایی بالاتری نسبت به روش‌های دیگر است.

جدول ۳، نتایج دقت طبقه‌بندی روش‌های مختلف انتخاب ویژگی با استفاده از روش k-نزدیکترین همسایه را نشان می‌دهد.

نتایج جدول ۳ نشان می‌دهد که روش انتخاب ویژگی پیشنهادی با استفاده از روش طبقه‌بندی k-نزدیکترین همسایه دارای دقت بالاتری نسبت به دیگر روش‌های انتخاب ویژگی است.

نتایج آزمایش‌ها حاکی از برتری روش انتخاب ویژگی پیشنهادی با هر دو روش طبقه‌بندی ماشین بردار پشتیبان و k-نزدیکترین همسایه نسبت به سایر روش‌های انتخاب ویژگی است. این نتایج نشان می‌دهد که استفاده از اطلاعات برچسب به همراه ساختار محلی داده‌های برچسب‌دار و بدون برچسب در محاسبه ماتریس پراکندگی پیشنهادی مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر، منجر به بهبود کارایی روش پیشنهادی می‌گردد و ویژگی‌های مناسب‌تری انتخاب می‌شوند.

۴-۳- مقایسه روش پیشنهادی و روش GS^3FS- $l_{2,1}$

به منظور مقایسه دقیق‌تر روش پیشنهادی SFSHF و روش قبلی پیشنهادی GS^3FS- $l_{2,1}$ مبتنی بر لاپلاسین گراف [۳۶] ، آزمایش‌هایی با استفاده از روش ارزیابی ارائه شده در [۳۶] انجام می‌شود و دقت این دو روش اندازه‌گیری می‌شود. در این آزمایش‌ها، همانند ارزیابی‌های انجام شده در [۳۶]، طبقه‌بندی با استفاده از تعداد یکسانی از ویژگی‌های انتخاب شده توسط روش انتخاب ویژگی انجام می‌شود و مقادیر بهینه پارامتر منظم‌سازی در روش پیشنهادی و روش GS^3FS- $l_{2,1}$ با استفاده از روش اعتبارسنجی متقاطع ۱۰-دسته‌ای و مجموعه {۰/۰،۰۰۱/۰،۰۱/۱۰۰۰،۱۰۰،۱۰،۱،۱} بر روی مجموعه داده‌های آموزشی تعیین می‌شوند. تعداد ویژگی‌های انتخاب شده برای مجموعه داده‌های دیابت، WBCD و WDBC همانند مرجع [۳۶] در نظر گرفته می‌شود. جدول ۴، نتایج ارزیابی روش پیشنهادی SFSHF و روش GS^3FS- $l_{2,1}$ با استفاده از تعداد یکسانی ویژگی را با روش‌های طبقه‌بندی ماشین بردار پشتیبان و k نزدیکترین همسایه (k=1 , k=5) نشان می‌دهد. اعداد داخل پرانتز در هر مجموعه داده، تعداد ویژگی‌های انتخاب شده را نشان می‌دهند.

جدول ۳- دقت طبقه‌بندی (میانگین± انحراف معیار) روش‌های مختلف انتخاب ویژگی با روش k-نزدیکترین همسایه

مجموعه داده	EGCFS	S^2DFS	SFSN	FSNM	S^2FS^2R	HFSL	GS^3FS- $l_{2,1}$	SFS	SFSHF
دیابت	۰/۷۰۲± ۰/۰۴۹	**۰/۷۱۵± ۰/۰۰۸**	۰/۶۹۰± ۰/۰۳۹	۰/۶۸۵± ۰/۰۴	۰/۶۹۹± ۰/۰۴	۰/۷۰۳± ۰/۰۴۰	۰/۷۰۱± ۰/۰۵۹	۰/۶۶۷± ۰/۰۸۰	**۰/۷۱۶± ۰/۰۵۵**
WBCD	۰/۹۵۱± ۰/۰۱۴	۰/۹۶۱± ۰/۰۱۴	۰/۹۴۶± ۰/۰۱۲	۰/۹۵۱± ۰/۰۱	۰/۹۵۴± ۰/۰۰۵	۰/۹۵۵± ۰/۰۱۵	۰/۹۶۰± ۰/۰۱۳	۰/۹۵۲± ۰/۰۲۴	**۰/۹۶۶± ۰/۰۱۹**
WDBC	۰/۹۲۴± ۰/۰۲۳	۰/۹۲۱± ۰/۰۱۶	۰/۹۲۹± ۰/۰۱۶	۰/۹۳۴± ۰/۰۱۱	۰/۹۳۱± ۰/۰۲۱	۰/۹۲۹± ۰/۰۱۱	۰/۹۳۸± ۰/۰۱۱	۰/۹۰۵± ۰/۰۳۲	**۰/۹۴۰± ۰/۰۰۸**
QSAR	۰/۷۸۸± ۰/۰۲۸	۰/۷۹۰± ۰/۰۱۳	۰/۷۶۳± ۰/۰۱۷	۰/۷۸۹± ۰/۰۱۹	۰/۷۹۲± ۰/۰۱۶	۰/۷۹۳± ۰/۰۲۱	۰/۸۰۹± ۰/۰۱۱	۰/۷۸۰± ۰/۰۰۵	**۰/۸۱۴± ۰/۰۲۰**
پارکینسون	۰/۸۰۸± ۰/۰۴۳	۰/۸۰۲± ۰/۰۲۰	۰/۷۸۰± ۰/۰۴۵	۰/۸۰۸± ۰/۰۲۶	۰/۸۰۲± ۰/۰۲۶	۰/۸۱۹± ۰/۰۲۰	۰/۸۰۸± ۰/۰۴۳	۰/۷۹۷± ۰/۰۲۹	**۰/۸۴۲± ۰/۰۴۳**

جدول۴- نتایج دقت طبقه‌بندی روش SFSHF و روش GS^{3}FS- $l_{2,1}$ با تعداد ویژگی‌های یکسان

مجموعه داده	روش انتخاب ویژگی	SVM	1NN	5NN
دیابت (۳)	SFSHF	۰/۶۷۷±۰/۰۰۲	۰/۶۰۱±۰/۰۲۳	۰/۶۵۲±۰/۰۱۸
	GS^{3}FS - $l_{2,1}$	**۰/۷۶۰±۰/۰۰۵**	۰/۶۵۸±۰/۰۴۳	۰/۷۰۳±۰/۰۵۱
دیابت (۵)	SFSHF	۰/۷۴۲±۰/۰۶۱	۰/۶۵۱±۰/۰۱۵	۰/۷۱۶±۰/۰۵۵
	GS^{3}FS - $l_{2,1}$	**۰/۷۵۱±۰/۰۵۶**	۰/۶۸۶±۰/۰۱۱	۰/۷۲۶±۰/۰۳۹
WBCD (۳)	SFSHF	**۰/۹۷۰±۰/۰۱۴**	۰/۹۱۸±۰/۹۲۲	۰/۹۴۹±۰/۰۱۰
	GS^{3}FS - $l_{2,1}$	۰/۹۶۵±۰/۰۱۵	۰/۹۲۲±۰/۰۱۱	۰/۹۴۷±۰/۰۰۲
WBCD (۵)	SFSHF	**۰/۹۷۲±۰/۰۰۹**	۰/۹۵۴±۰/۰۲۷	**۰/۹۷۱±۰/۰۰۶**
	GS^{3}FS - $l_{2,1}$	۰/۹۶۸±۰/۰۱۳	۰/۹۵۱±۰/۰۲۹	۰/۹۶۳±۰/۰۱۶
WDBC (۱۰)	SFSHF	**۰/۹۴۷±۰/۰۰۶**	۰/۹۱۲±۰/۰۲۱	۰/۹۴۰±۰/۰۰۸
	GS^{3}FS - $l_{2,1}$	۰/۹۴۰±۰/۰۲۳	۰/۹۰۵±۰/۰۲۱	۰/۹۳۶±۰/۰۰۶
WDBC (۱۵)	SFSHF	**۰/۹۵۳±۰/۰۱۴**	۰/۹۴۸±۰/۰۰۵	۰/۹۴۸±۰/۰۱۴
	GS^{3}FS - $l_{2,1}$	۰/۹۴۷±۰/۰۲۷	۰/۹۲۴±۰/۰۱۵	۰/۹۴۸±۰/۰۰۹
QSAR (۵)	SFSHF	**۰/۸۰۰۲±۰/۰۲۵**	۰/۷۵۸±۰/۰۰۵	۰/۷۷۵±۰/۰۳۲
	GS^{3}FS - $l_{2,1}$	۰/۷۸۹±۰/۰۲۳	۰/۷۴۲±۰/۰۴۷	۰/۷۵۴±۰/۰۱۵
QSAR (۱۰)	SFSHF	**۰/۸۳۴±۰/۰۰۹**	۰/۷۹۲±۰/۰۳۳	۰/۸۰۰±۰/۰۲۶
	GS^{3}FS - $l_{2,1}$	۰/۸۲۳±۰/۰۲۳	۰/۷۸۰±۰/۰۱۰	۰/۷۹۱±۰/۰۱۸
پارکینسون (۱۰)	SFSHF	۰/۸۰۲±۰/۰۸۵	۰/۸۰۸±۰/۰۲۶	**۰/۸۱۹±۰/۰۴۳**
	GS^{3}FS - $l_{2,1}$	۰/۸۰۲±۰/۰۲۰	۰/۷۸۰±۰/۰۴۵	۰/۷۷۴±۰/۰۷۱
پارکینسون (۱۵)	SFSHF	**۰/۸۷۶±۰/۰۲۰**	۰/۸۴۸±۰/۰۳۴	۰/۸۲۹±۰/۰۲۶
	GS^{3}FS - $l_{2,1}$	۰/۸۵۹±۰/۰۲۰	۰/۸۲۵±۰/۰۱۰	۰/۸۱۴±۰/۰۱۷

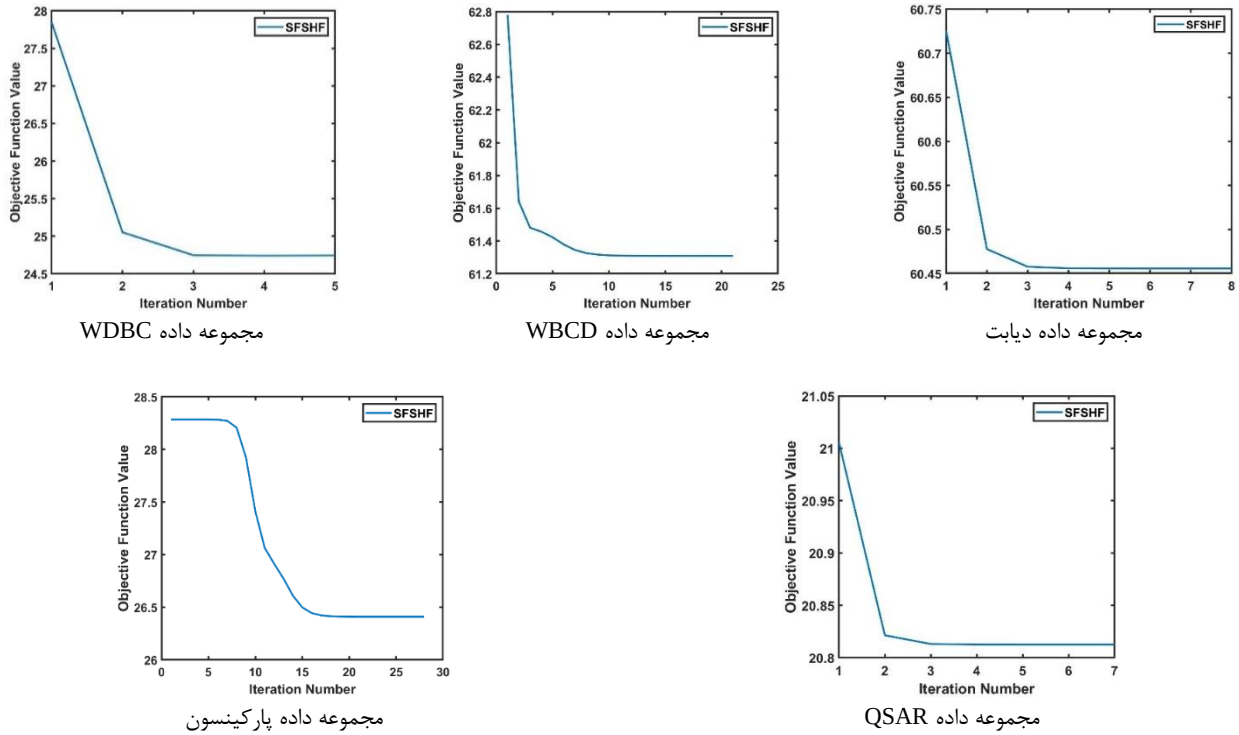
همان‌گونه که در جدول ۴ نشان داده شده است، روش پیشنهادی SFSHF با در نظر گرفتن تعداد یکسانی از ویژگی‌ها بر روی تمام مجموعه داده‌ها به جز مجموعه داده دیابت، دارای دقت بالاتری نسبت به روش GS^{3}FS- $l_{2,1}$ است.

۴-۴- مطالعه همگرایی

در این بخش، به منظور بررسی عملی همگرایی الگوریتم ۱، آزمایش‌هایی انجام می‌شود. در آزمایش‌ها، پارامتر λ در رابطه (۲) به ۱۰ مقداردهی می‌شود. شکل ۲ منحنی‌های همگرایی مقادیر تابع هدف رابطه (۲) مبتنی بر ماتریس هسین و آنالیز تشخیصی فیشر را با استفاده از الگوریتم ۱ بر روی پنج مجموعه داده نشان می‌دهد. همان‌گونه که در این شکل مشخص است، مقادیر تابع هدف با استفاده از الگوریتم ۱ به سرعت همگرا می‌شوند که نشان‌دهنده کارایی و موثر بودن الگوریتم ۱ بر روی تابع هدف پیشنهادی است.

۵- نتیجه‌گیری

در این مقاله، یک روش انتخاب ویژگی نیمه‌نظارتی تُنک مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر پیشنهاد شد تا با استفاده از نُرم- $l_{2,1}$، اطلاعات برچسب داده‌های برچسب‌دار و ساختار محلی داده‌های برچسب‌دار و بدون برچسب، مناسب‌ترین ویژگی‌ها را انتخاب نماید. روش پیشنهادی از یک ماتریس پراکندگی نیمه نظارتی استفاده می‌نماید که این ماتریس ترکیب دو ماتریس پراکندگی نظارتی مبتنی بر آنالیز تشخیصی فیشر و ماتریس پراکندگی غیرنظارتی مبتنی بر منظم‌سازی هسین است. منظم‌سازی هسین توزیع و ساختار محلی داده‌ها را به خوبی حفظ می‌کند و منجر به قدرت برون‌یابی خوبی برای افزایش کارایی روش انتخاب ویژگی پیشنهادی می‌گردد. برای حل در روش پیشنهادی مبتنی بر منظم‌سازی هسین و آنالیز تشخیصی فیشر، الگوریتمی با رویکرد تکراری استفاده شد و همگرایی آن به صورت تئوری و عملی اثبات گردید.



مجموعه داده دیابت — مجموعه داده WBCD — مجموعه داده WDBC

مجموعه داده QSAR — مجموعه داده پارکینسون

شکل ۲- منحنی‌های همگرایی الگوریتم ۱ مطابق با مقادیر تابع هدف در رابطه (۲)

کارایی روش پیشنهادی با استفاده از پنج مجموعه داده و دو روش طبقه‌بندی ماشین بردار پشتیبان و k-نزدیکترین همسایه با دیگر روش‌های انتخاب ویژگی مقایسه گردید. نتایج آزمایش‌ها، برتری روش پیشنهادی در مقایسه با سایر روش‌های انتخاب ویژگی را نشان دادند که حاکی از برتری و کارایی منظم‌سازی هسین در محاسبه ماتریس پراکندگی پیشنهادی و حفظ ساختار محلی داده‌ها است.

مراجع

[1] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Computers and Electrical Engineering. 40 (2014) 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

[2] Y. Hu, Y. Zhang, D. Gong, Multiobjective Particle Swarm Optimization for Feature Selection With Fuzzy Cost, IEEE TRANSACTIONS ON CYBERNETICS. 51 (2021) 874–888. https://doi.org/10.1109/TCYB.2020.3015756.

[3] G. Dhiman, D. Oliva, A. Kaur, K.K. Singh, S. Vimal, A. Sharma, K. Cengiz, BEPO: A novel binary emperor penguin optimizer for automatic feature selection, Knowledge-Based Systems. 211 (2021). https://doi.org/10.1016/j.knosys.2020.106560.

[4] J. Wang, H. Zhang, J. Wang, Y. Pu, N.R. Pal, Feature Selection Using a Neural Network With Group Lasso Regularization and Controlled Redundancy, IEEE Transactions on Neural Networks and Learning Systems. 32 (2021) 1110–1123.

[۵] ح. بیاتی، م. دولتشاهی، م. پنیری، انتخاب ویژگی چندبرچسبی با استفاده از الگوریتم بهینه ساز جمعیت رقابتی، مجله علمی رایانش نرم و فناوری اطلاعات. ۹ (۲۰۲۰) ۵۶–۶۹ .

[۶] س. حیدری مقدم بجستانی، س. شعرباف تبریزی ، ع. قاضی خانی ، ارائه‌ی یک روش انتخاب ویژگی جدید مبتنی بر بهینه سازی ازدحام ذرات با استفاده از به‌روزرسانی فازی، مجله مهندسی برق دانشگاه تبریز. ۵۰ (۲۰۲۱) ۱۵۶۷–۱۵۵۷ .

[7] W. Zhong, X. Chen, F. Nie, J. Zhexue, Adaptive discriminant analysis for semi-supervised feature selection, Information Sciences. 566 (2021) 178–194. https://doi.org/10.1016/j.ins.2021.02.035.

[8] M. Tubishat, S. Ja, M. Alswaitti, S. Mirjalili, Dynamic Salp Swarm Algorithm for Feature Selection, Expert Systems with Applications. 164 (2021) 113873. https://doi.org/https://doi.org/10.1016/j.eswa.2020.113873.

[9] G. ROFFO, S. Melzi, U. Castellani, A. Vinciarelli, M. Cristani, Infinite Feature Selection: a Graph-based Feature Filtering Approach, IEEE Transactions on Pattern Analysis and Machine Intelligence. (2020). https://doi.org/10.1109/TPAMI.2020.3002843.

[۱۰] م. رحمانی نیا، پ. مرادی، م. جلیلی ، یک راهکار انتخاب ویژگی چندهدفه بر اساس اطلاعات متقابل شرطی و نظریه مجموعه پارتو، مجله مهندسی برق دانشگاه تبریز. ۵۰ (۲۰۲۰) ۱۲۳۷–۱۲۲۵ .

[11] M. Sharifnezhad, M. Rahmani, A. Professor, H. Ghaffarian, A Distributed Minimum Redundancy Maximum Relevance Feature Selection Approach, Tabriz Journal of Electrical Engineering (TJEE). 51 (2021) 286–293.

[12] Razieh Sheikhpour; Mehdi Agha Sarrama; Sajjad Gharaghani; Mohammad Ali Zare Chahookia, R. Sheikhpour, M.A.M.A. Sarram, S. Gharaghani, M.A.Z.M.A.Z. Chahooki, A Survey on semi-supervised feature selection methods, Pattern Recognition. 64 (2017) 141–158. https://doi.org/10.1016/j.patcog.2016.11.003.

[13] T. Bhadra, S. Bandyopadhyay, Supervised feature selection using integration of densest subgraph finding with floating forward – backward search, Information Sciences. 566 (2021) 1–18. https://doi.org/10.1016/j.ins.2021.02.034.

[14] R. Zhang, H. Zhang, X. Li, S. Yang, Unsupervised Feature Selection With Extended OLSDA via Embedding Nonnegative Manifold Structure, IEEE Transactions on Neural Networks and Learning Systems. (2020) 1–7.

[15] Q. Pang, L. Zhang, Semi-supervised neighborhood discrimination index for feature selection, Knowledge-Based Systems. 204 (2020) 106224. https://doi.org/10.1016/j.knosys.2020.106224.

[16] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Adv Neural Inf Process Syst, 2005: pp. 507–514.

[17] G. Doquire, M. Verleysen, A graph Laplacian based approach to semi-supervised feature selection for regression problems, Neurocomputing. 121 (2013) 5–13.

[18] J. Zhao, K. Lu, X. He, Locality sensitive semi-supervised feature selection, Neurocomputing. 71 (2008) 1842–1849. https://doi.org/10.1016/j.neucom.2007.06.014.

[19] Z. Ma, F. Nie, Y. Yang, J.R.R. Uijlings, N. Sebe, S. Member, A.G. Hauptmann, Discriminating joint feature analysis for multimedia data understanding, IEEE TRANSACTIONS ON MULTIMEDIA. 14 (2012) 1662–1672.

[20] C. Shi, Q. Ruan, G. An, Sparse feature selection based on graph Laplacian for web image annotation, Image and Vision Computing. 32 (2014) 189–201. https://doi.org/10.1016/j.imavis.2013.12.013.

[21] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, S. Member, Semisupervised feature selection via spline regression for video semantic recognition, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,. 26 (2015) 252–264.

[22] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological). 58 (1996) 267–288.

[23] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via ℓq-minimization for 0< q<1, Applied and Computational Harmonic Analysis. 26 (2009) 395–407.

[24] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Processing Letters. 14 (2007) 707–710. https://doi.org/10.1109/LSP.2007.898300.

[25] F. Nie, H. Huang, X. Cai, C.H. Ding, Efficient and robust feature selection via joint ℓ2, 1-norms minimization, in: Adv Neural Inf Process Syst, 2010: pp. 1813–1821.

[26] L. Wang, S. Chen, l2,p-matrix norm and its application in feature selection, ArXiv Preprint ArXiv:1303.3987. (2013).

[27] C. Shi, Q. Ruan, S. Member, G. An, R. Zhao, Hessian semi-supervised sparse feature selection based on L21/2-matrix norm, IEEE Transactions on Multimedia. 17 (2015) 16–28.

[28] C.M. Bishop, Neural networks for pattern recognition, Oxford University Press, 1995.

[29] Q. Gu, Z. Li, J. Han, Generalized Fisher Score for Feature Selection, CoRR. abs/1202.3 (2012).

[30] M. Yang, Y. Chen, G. Ji, Semi_fisher score : a semi-supervised method for feature selection, in: International Conference on Machine Learning and Cybernetics, 2010: pp. 527–532.

[31] S. Lv, H. Jiang, L. Zhao, D. Wang, M. Fan, Manifold based fisher method for semi-supervised feature selection, in: 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2013: pp. 664–668.

[32] L. Chen, R. Huang, W. Huang, Graph-based semi-supervised weighted band selection for classification of hyperspectral data, in: Audio Language and Image Processing (ICALIP), 2010 International Conference On, IEEE, 2010: pp. 1123–1126.

[33] W. Yang, C. Hou, Y. Wu, A semi-supervised method for feature selection, 2011 International Conference on Computational and Information Sciences. (2011) 329–332. https://doi.org/10.1109/ICCIS.2011.54.

[34] Y. Liu, F. Nie, J. Wu, L. Chen, Efficient semi-supervised feature selection with noise insensitive trace ratio criterion, Neurocomputing. 105 (2013) 12–18. https://doi.org/10.1016/j.neucom.2012.05.031.

[35] Y. Liu, F. Nie, J. Wu, L. Chen, Semi-supervised feature selection based on label propagation and subset selection, in: Computer and Information Application (ICCIA), 2010 International Conference On, IEEE, 2010: pp. 293–296.

[36] R. Sheikhpour, M.A. Sarram, S. Gharaghani, M.A.Z. Chahooki, A robust graph-based semi-supervised sparse feature selection method, Information Sciences. 531 (2020) 13–30. https://doi.org/10.1016/j.ins.2020.03.094.

[37] R. Sheikhpour, M.A. Sarram, E. Sheikhpour, Semi-supervised sparse feature selection via graph Laplacian based scatter matrix for regression problems, Information Sciences. 468 (2018) 14–28. https://doi.org/10.1016/j.ins.2018.08.035.

[38] X. Li, Y. Zhang, R. Zhang, Semisupervised Feature Selection via Generalized Uncorrelated Constraint and Manifold Embedding, IEEE Transactions on Neural Networks and Learning Systems. (2021). https://doi.org/10.1109/TNNLS.2021.3069038.

[39] R. Sheikhpour, M.A. Sarram, S. Gharaghani, M.A.Z. Chahooki, Feature selection based on graph Laplacian by using compounds with known and unknown activities, Journal of Chemometrics. (2017). https://doi.org/10.1002/cem.2899.

[40] K.I. Kim, F. Steinke, M. Hein, Semi-supervised Regression using Hessian Energy with an Application to Semi-supervised Dimensionality Reduction, in: Advances in Neural Information Processing Systems (NIPS). MPI for Biological Cybernetics, Germany, 2010: pp. 979–987.

[41] R. Zhang, Y. Zhang, X. Li, Unsupervised Feature Selection via Adaptive Graph Learning and Constraint, IEEE Transactions on Neural Networks and Learning Systems. (2020). https://doi.org/10.1109/TNNLS.2020.3042330.

[42] Z. Wang, F. Nie, L. Tian, R. Wang, X. Li, Discriminative Feature Selection via A Structured Sparse Subspace Learning Module, in: IJCAI, 2020: pp. 3009–3015.