Journal of Iranian Association of Electrical and Electronics Engineers
Vol. 22 No. 2, 2025
DOI: 10.61186/jiaeee.22.2.93

Analyzing of Positive and Negative Prototypes based on the ±ED-WTA Method

Mahsa Fazaeli-Javan[1] , Reza Monsefi[2] , Kamaledin Ghiasi-Shirazi[3]

[1] Ph.D. candidate, Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
m.fazaelijavan@mail.um.ac.ir
[2] Professor, Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
monsefi@um.ac.ir
[3]Associate Professor, Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
k.ghiasi@um.ac.ir

Abstract

The recently introduced method, ±ED-WTA, for each output neuron of a class in the last layer of neural networks, obtain a pair of positive and negative prototypes. The functionality of a neuron is explained based on the difference in the Euclidean distance of a sample from these two prototypes. An astonishing point in this method is the great similarity of positive and negative prototypes for each neuron at the end of training. The authors of [2] have claimed that the reason for this extreme similarity is the formation of the negative prototype of a class with samples of other classes, which are very similar to the positive prototype of that class. In this paper, it is demonstrated that not only does the negative prototype get close to the positive prototype of a class, but a positive prototype is also formed by samples that are far from the center of that class. The new finding about this great similarity shows that each neuron in the softmax layer, like SVMs, makes decisions based on the near-boundary samples. The theoretical analysis is examined in detail and experimental results on MNIST, FERET, and Fashion-MNIST show the correctness of the claims made.

Keywords: Prototype-based learning, Positive prototype, Negative prototype, Boundary samples, Support vector machine

Article Type: Research

Received: 11. 02. 2023

Revised: 05. 06. 2024

Accepted: 09. 19. 2024

Corresponding author: R. Monsefi

Corresponding author's address: Kalantari Ave. Comp. Group, Eng. Dep., Ferdowsi Uni. Mashhad, Iran

1. Motivation of the work

In many prototype-based classifications like LVQ, NCM, and PROTONET [3], only the positive prototype—an average of class samples—is utilized. ±ED-WTA introduces the concept of incorporating negative prototypes alongside positive ones for each center in the Softmax layer, creating a new interpretable classification model. The positive prototype represents samples belonging to the class of that neuron, while the negative prototype represents samples that the neuron misclassifies. The method exhibits a significant similarity between positive and negative prototypes after training, making them difficult to distinguish. The proposed method have been analyzed ±ED-WTA to find the exact definition of positive and negative prototypes and reveal the reasons for the great similarity between them.

2. Contributions

The explanation in [2] for the high similarity between positive and negative prototypes is that negative prototypes are derived from samples mistakenly classified by neurons of a class due to their similarity to positive prototypes. However, this article argues that this explanation is insufficient. It explains that not only do negative prototypes shift closer to positive ones in similarity, but positive prototypes also move away from the class center and approach the boundaries between classes. This article explains how samples with less similarity to class centers significantly influence prototype formation. The main contributions are as follows:

- This paper presents a comprehensive interpretation of positive and negative prototypes based on the effective samples that contribute to them.
- The relationship between how neurons classify samples in the last layer of neural networks and SVM classifier has been revealed.

3. Procedures

In the proposed method section, first the gradient descent optimization of prototypes in the ±ED-WTA method has been analyzed. It shows that prototypes, represented as the weighted average of samples, are influenced by samples with larger gradients during updates.

Then, it is demonstrated that effective samples which have a larger gradient in updating the positive prototype of their class, also have a significant gradient in updating the negative prototypes of the other classes.

The gradient magnitude of samples and histogram analysis indicate that the effective training data contributing to positive and negative prototypes reduces in the final training epochs.

It also shows that these limited data are almost close to the boundary of the classes and have less similarity to their class center.

The final part of the proposed method section analyzes the relationship between support vector machine (SVM) and the ±ED-WTA method. It demonstrates that interpreting neuron functionality through the ±ED-WTA method shows that support vectors in SVM, which are boundary samples, also play a significant role in neuron decision-making.

4. Findings

The findings of this article have been demonstrated on MNIST, FERET, and Fashion-MNIST datasets.

To investigate the similarity of positive and negative prototypes, based on the definition of the article [2], a method called Mean-PNPL has been implemented, in each step of training, the positive prototype is the average of the correctly classified samples, and the negative prototype is the average of samples that are wrongly classified by an output neuron. In fact, in Mean-PNPL method, the interpretation given by [2] about the positive and the negative prototypes and the reason for their great similarity has been included. The results show that the positive prototypes are slightly different from the average of the class samples. The positive and negative prototypes of a center are more distinguishable, and the accuracy of the model is lower compared to the accuracy of the ±ED-WTA.

The relationship between SVM and ±ED-WTA was demonstrated by examining the multi-prototype SVM on the above datasets. The results indicate that samples selected as support vectors also significantly impact the formation of both positive and negative prototypes.

5. Conclusion

The ±ED-WTA method interprets the functionality of a neuron based on the squared difference of the Euclidean distance from the positive prototype and the negative prototype. This article analyzes the gradient descent optimization of prototypes in the ±ED-WTA. It presents a complete definition of positive and negative prototypes based on the effective samples in their formation. The new finding shows that positive and negative prototypes are made from near-boundary samples. The new theory implicitly states that neurons in the competitive layer of the neural network, similar to SVM, make decisions based on boundary data.

نشریه مهندسی برق و الکترونیک ایران

نشریه مهندسی برق و الکترونیک ایران، سال بیست و دوم، شماره دوم، تابستان ۱۴۰۴، صفحات ۹۳-۱۰۷

DOI: 10.61186/jiaeee.22.2.93

تحلیل الگوهای مثبت و منفی در ED-WTA±

مهسا فضائلی جوان[۱]، رضا منصفی[۲] ، سید کمال‌الدین غیاثی شیرازی[۳]

۱- دانشجوی دکتری- بخش مهندسی کامپیوتر- دانشکده مهندسی ـ دانشگاه فردوسی مشهدـ مشهدـ ایران
m.fazaelijavan@mail.um.ac.ir

۲- استاد- بخش مهندسی کامپیوتر-دانشکده مهندسی -دانشگاه فردوسی مشهد- مشهد- ایران
monsefi@um.ac.ir

۳- دانشیار-بخش مهندسی کامپیوتر-دانشکده مهندسی -دانشگاه فردوسی مشهد- مشهد- ایران
k.ghiasi@um.ac.ir

چکیده: روش ED-WTA± یک روش جدید یادگیری مبتنی بر الگو برای شبکه‌های عصبی مرسوم است که متناظر با هر نورون خروجی یک جفت الگوی مثبت و منفی به دست می‌آورد. الگوی مثبت، نمایان‌گر نمونه‌هایی است که نورون به درستی در مورد آن‌ها تصمیم می‌گیرد و الگوی منفی نماینده‌ی نمونه‌هایی است که نورون به اشتباه به ازای آن‌ها برنده شده است. ED-WTA± برای تفسیر عملکرد نورون در لایه‌ی بیشینه‌ی نرم، اختلاف مربع فاصله‌ی هر داده از الگوی مثبت و الگوی منفی را لحاظ می‌کند. نکته‌ی جالب توجه در این روش این است که در انتهای آموزش، الگوی مثبت و منفی هر نورون شباهت بسیار زیادی با یکدیگر پیدا می‌کنند. در این مقاله، با تحلیل چگونگی تشکیل الگوها در روش ED-WTA±، نشان داده می‌شود، دلیل شباهت بسیار زیاد الگوهای مثبت و منفی، تأثیرگذاری آن‌ها بر یکدیگر از طریق داده‌های نزدیک به مرز کلاس‌ها می‌باشد. همچنین با تعیین نمونه‌های مؤثر در شکل‌گیری الگوها، مشاهده می‌شود، همان‌طور که نمونه‌های مرزی در طبقه‌بند ماشین بردار پشتیبان مرز تصمیم‌گیری را تعیین می‌کنند، در شبکه‌ی عصبی نیز نورون‌های لایه‌ی آخر بر اساس نمونه‌های مرزی طبقه‌بندی را انجام می‌دهند. آزمایشات بر روی مجموعه دادگان MNIST، FERET و Fashion-MNIST درستی ادعاهای مطرح شده را نشان می‌دهد.

کلمات کلیدی: یادگیری مبتنی بر الگو، الگوی مثبت، الگوی منفی، داده‌های مرزی، ماشین بردار پشتیبان

نوع مقاله: پژوهشی

دریافت: ۱۴۰۲/۰۸/۱۱

بازنگری: ۱۴۰۳/۰۳/۱۶

پذیرش: ۱۴۰۳/۰۶/۲۹

نام نویسنده‌ی مسئول: دکتر رضا منصفی

نشانی نویسنده‌ی مسئول: ایران - مشهد - بلوار وکیل آباد - بلوار باهنر - دانشگاه فردوسی - دانشکده‌ی مهندسی- بخش کامپیوتر

۱- مقدمه

در روش‌های یادگیری مبتنی بر الگو، برای هر کلاس یک یا چند الگو تعریف می‌شود که این الگوها یا یک نمونه‌ی مشخص از هر کلاس هستند و یا ترکیب محدبی از کل دادگان یک کلاس می‌باشند [۴]. سپس برچسب یک نمونه‌ی تست با توجه به فاصله‌ای که تا این الگوها دارد تعیین می‌شود. روش‌های طبقه‌بندی مبتنی بر الگو تفسیرپذیری بهتری از روش‌هایی نظیر ماشین بردار پشتیبان[۱] و شبکه‌های عصبی پیچشی[۲] که مبتنی بر ابرصفحه هستند ارائه می‌دهند. همچنین برای کاربردهایی نظیر طبقه‌بندی اندک-نمونه[۳] که تعداد نمونه‌های هر کلاس اندک هستند، انتخاب یک نماینده برای هر کلاس، استراتژی مناسبی برای تعمیم مدل بر روی داده‌های تست است [۳].

در بسیاری از تحقیقات در زمینه‌ی طبقه‌بندی مبتنی بر الگو نظیر PROTONET [3]، با الهام از روش LVQ[۴] فقط از الگوی مثبت استفاده شده است که میانگین نمونه‌های یک کلاس است. در ±ED-WTA[۵] ایده‌ی لحاظ کردن الگوهای منفی علاوه بر الگوهای مثبت برای هر نورون که نمایان‌گر یک زیر کلاس می‌باشد را به منظور ارائه‌ی مدل جدیدی از طبقه‌بندی تفسیرپذیر ارائه داده است.

±ED-WTA یک مدل طبقه‌بندی مبتنی بر الگو در شبکه‌های عصبی winner-take-all (WTA) است که با یک مدل تفسیرپذیر مبتنی بر الگو عملکرد نورون در لایه‌ی بیشینه‌ی نرم[۶] شبکه‌ی عصبی تک-لایه را مدل‌سازی می‌نماید. در ±ED-WTA هر نورون با دو الگو مدل می‌شود: الگوی مثبت که نمایان‌گر نمونه‌هایی است که به درستی با آن نورون در مورد کلاس آن‌ها تصمیم‌گیری می‌شود و الگوی منفی که نمایان‌گر نمونه‌هایی است که به اشتباه، آن نورون به ازای آن‌ها برنده شده است. در این روش با الهام از مقاله‌ی [۲۸] برای هر کلاس بیش از یک نورون در لایه‌ی بیشینه‌ی نرم در نظر گرفته شده است. با توجه به اختلاف مربع فاصله‌ی یک نمونه تا الگوی مثبت و فاصله‌اش تا الگوی منفی نورون، مقدار فعالیت آن نورون محاسبه می‌گردد.

در مدل ±ED-WTA عملکرد نورون بر اساس اختلاف مربع فاصله از الگوی مثبت و الگوی منفی مدل می‌شود. در هنگام آموزش، ضریب مثبتی از نمونه‌ی آموزشی به وزن الگوهای مثبت از کلاس موافق اضافه می‌گردد و ضریب مثبت دیگری از این نمونه به وزن الگوهای منفی کلاس‌های دیگر اضافه می‌شود. به این ترتیب با این نحوه‌ی به‌روزرسانی، ماهیت الگو بودن وزن‌ها و شباهت آن‌ها به نمونه-های کلاسی حفظ می‌گردد.

نکته اساسی و قابل تأملی که در این روش وجود دارد شباهت بسیار زیاد الگوهای مثبت و منفی پس از پایان آموزش است که به سختی قابل تمایز است. دلیلی که در [۲] برای این نکته آمده است، ساخته شدن الگوهای منفی از روی نمونه‌هایی است که با نورون یک کلاس به اشتباه برنده می‌شوند زیرا بسیار شبیه به الگوهای مثبت کلاس هستند. در این مقاله با تحلیل روابط مربوط به به‌روزرسانی الگوها شرح داده می‌شود که این دلیل کافی نمی‌باشد و تنها الگوهای منفی نیستند که به تدریج از نظر شباهت به سمت الگوهای مثبت میل می‌کنند بلکه الگوهای مثبت نیز در حال تغییر و فاصله گرفتن از مراکز زیرکلاس هستند. دلیل مهم این رویداد که در این مقاله به اثبات آن پرداخته می‌شود آن است که نمونه‌های مثبتی از کلاس که شباهت کمتری به مراکز زیرکلاس دارند بر روی الگوی مثبت تأثیر بیشتری دارند و نمونه‌هایی که نزدیک به مرکز زیرکلاس هستند و با درجه اطمینان بالا به درستی دسته‌بندی می‌شوند بر روی الگوهای مثبت تأثیر بسیار کمی دارند. به این ترتیب الگوهای مثبت و منفی فاصله‌ی بسیار کمی با هم پیدا می‌کنند.

هر چند در تحقیقات دیگری که در زمینه‌ی طبقه‌بندی مبتنی بر الگو است نیز اصطلاح الگوی منفی بکار رفته است [۵ , ۶]. اما در این مقالات منظور الگوی مثبتی از کلاس مخالف است که نزدیک‌ترین فاصله را تا کلاس مورد نظر دارد. در این روش‌ها با اضافه کردن قید بیشینه کردن فاصله تا الگوی منفی و کمینه کردن فاصله تا الگوی مثبت در تابع هدف، جداپذیری بیشتر کلاس‌ها تأمین می‌شود. اما در این تحقیقات هر کلاس در واقع یک یا چند الگوی مثبت دارد و مدل طبقه‌بندی همان مدل طبقه‌بندی مبتنی بر الگوی مثبت است.

همچنین در این مقاله نشان داده می شود، در ±ED-WTA در هنگام آموزش، نمونه‌هایی مؤثرتر در تشکیل الگوها هستند که شباهت کمتری به نمونه‌های کلاس خود دارند و یا به اصطلاح نمونه‌های سخت[۷] هستند. نام این پدیده، کاوش سخت-آگاه الگو[۸] نامیده می‌شود. با توجه به کم بودن نسبت نمونه‌های سخت به کل دادگان آموزشی، در انتهای آموزش، نمونه‌های محدودی هستند که در تشکیل الگوهای مثبت و منفی نقش دارند.

اهمیت دادن به نمونه‌های سخت در روش‌های طبقه‌بندی دیگر نیز که مرز تصمیم‌گیری را دادگان کلاس‌ها تعیین می‌کنند، مورد توجه بوده است. از جمله می‌توان به طبقه‌بند پرکاربرد SVM[۹] اشاره نمود. در SVM نیز نمونه‌های محدودی به عنوان بردارهای پشتیبان انتخاب می‌شوند که نزدیک به مرز تصمیم قرار گرفته‌اند و همین دادگان محدود ابرصفحه جداکننده کلاس‌ها را می‌سازند. در تحقیقات اخیر نیز اهمیت نمونه‌های مرزی در بالا بردن دقت طبقه‌بندی شبکه عصبی عمیق مورد تأکید و توجه قرار گرفته است [۷]. به این ترتیب کشف ارتباط طبقه‌بند مبتنی بر شبکه عصبی و SVM از نظر استراتژی انتخاب نمونه‌های مؤثر در تصمیم‌گیری، نوآوری دیگری است که در این مقاله به آن پرداخته می‌شود.

به طور کلی در پژوهش حاضر، نوآوری‌های ذیل حاصل شده است:

- تفسیر کامل از مفهوم الگوی مثبت و الگوی منفی با تکیه بر نمونه‌های مؤثر بر شکل‌گیری آن‌ها ارائه می‌گردد.
- با نشان دادن اهمیت نمونه‌های مرزی در شکل‌گیری ابرصفحه‌ی تصمیم‌گیری در لایه‌ی بیشینه‌ی نرم شبکه عصبی، ارتباط بین نحوه‌ی طبقه‌بندی نمونه‌ها توسط نورون‌ها و طبقه‌بندی SVM، در این مقاله، آشکار می‌گردد.

در بخش ۲ مروری بر روش‌های طبقه‌بندی مبتنی بر الگو انجام می‌شود. در بخش سوم روش ED-WTA± مرور خواهد شد. در بخش ۴ به واکاوی و تحلیل الگوهای مثبت و منفی و دلایل شباهت آن‌ها پرداخته می‌شود. سپس در این بخش آنالیز و مقایسه روش ED-WTA± و SVM تشریح می‌گردد. در بخش ۵ نتایج تجربی برای اثبات ادعاهایی که در این مقاله آمده است، آورده شده است. بخش ۶ نیز نتیجه‌گیری حاصل از پژوهش انجام شده ذکر گردیده است.

۲- تاریخچه

در این بخش ابتدا مروری بر تحقیقات مهمی می‌شود که در زمینه‌ی طبقه‌بندی مبتنی بر الگو انجام شده است. سپس با توجه به این‌که در این مقاله اهمیت نمونه‌های سخت در بالابردن دقت طبقه‌بندی نشان داده می‌شود، تحقیقاتی مرور می‌گردد که اهمیت این‌گونه نمونه‌ها را در طبقه‌بندی نشان داده‌اند.

برخی تحقیقات به منظور ارائه‌ی یک مدل یادگیری تفسیرپذیر، طبقه‌بندی مبتنی بر الگو ارائه داده‌اند تا بر اساس شباهت، دلیل دسته‌بندی هر نمونه در هر کلاس را توضیح دهند.

در این میان مقالاتی مانند [۹, ۸] هستند که با انتخاب الگوها از میان نمونه‌ها، علاوه بر هدف تفسیرپذیری، مجموعه دادگان را نیز به صورت خلاصه توسط الگوها نمایش می‌دهند. در [۸] تعداد حداقلی از نمونه‌ها را از کل مجموعه داده‌ها به عنوان الگو انتخاب می‌نماید به طوری که هر الگو به تعداد حداکثری از داده‌های کلاس خود نزدیک باشد و کم‌ترین تعداد داده‌ها از کلاس‌های دیگر در اطراف آن قرار داشته باشد. protoAttend [۹] نیز مانند روش [۸] از بین نمونه‌های ورودی با استفاده از مکانیزم توجه[۱۰]، تعدادی حداقل نمونه به عنوان الگو برای ارائه‌ی یک طبقه‌بند تفسیرپذیر انتخاب می‌نماید. معماری این روش شامل یک ماژول توجه وابستگی[۱۱] برای یادگیری میزان ارتباط یک نمونه با یک الگو است. در تابع زیان فاز استنتاج، درجه اطمینان از برچسب پیش‌بینی شده برای هر کلاس نیز مشخص می‌شود تا به این ترتیب علاوه بر مشخص کردن میزان اعتماد به نتایج، از بین نمونه‌ها الگوی مناسب‌تری نیز برای هر کلاس انتخاب گردد.

روش‌های یادگیری مبتنی بر الگویی نیز هستند که به دنبال تفسیرپذیر نمودن مدل طبقه‌بندی مبتنی بر شبکه‌ی عصبی عمیق می‌باشند. این تحقیقات به طور معمول الگو را ازمیان نمونه‌های مجموعه داده‌ی آموزشی انتخاب می‌نمایند تا مدل توضیح پذیر بهتری را ارائه دهند. یکی از این روش‌ها در [۴] ارائه شده است. این روش، یک مدل طبقه‌بندی بر اساس معماری شبکه عصبی عمیق ارائه می‌دهد که علت پیش‌بینی نتیجه شده برای کلاس هر داده به خوبی قابل بیان[۱۲] باشد. بعد از لایه‌ی انکدر برای استخراج ویژگی‌ها از شبکه عصبی عمیق، یک لایه الگوها هستند که فاصله‌ی هر الگو با بردار ویژگی استخراج شده محاسبه می‌گردد و مجموع وزن‌دار فاصله‌ی یک نمونه تا تمام الگوها به لایه‌ی بیشینه‌ی نرم داده می‌شود که به تعداد کلاس‌ها نورون دارد. همچنین یک دیکدر نیز برای بازسازی الگوها به فضای ورودی نیز در نظر گرفته شده است تا نزدیکی هر الگو در فضای عمیق به یکی از داده‌های کلاس خود در فضای ورودی نیز مورد بررسی قرار گیرد. روش ProtoPNet [۵] برای ارائه یک مدل تفسیرپذیر مبتنی بر الگو، در مقایسه با روش [۴] بخش‌های مختلف یک تصویر نمونه برای هر کلاس را به عنوان الگو انتخاب می‌نماید. قسمت‌های مختلف یک تصویر نمونه با الگوهای استخراج شده برای هر کلاس مقایسه می‌شود و بر اساس مجموع امتیازهای شباهتی که برای بخش‌های مختلف تصویر ورودی و نورون‌های لایه‌ی الگو بدست آمده است در مورد کلاس نمونه‌ی ورودی تصمیم‌گیری انجام می‌شود. در این روش نیز امتیازهای شباهت بر اساس فاصله‌ی اقلیدسی محاسبه می‌شود.

علاوه بر تحقیقات ذکر شده که الگوها را از میان نمونه‌ها انتخاب می‌کنند، برخی تحقیقات ترکیبی از نمونه‌های یک کلاس را به عنوان الگو در نظر می‌گیرند. این تحقیقات به ویژه در کاربردهایی نظیر یادگیری اندک-نمونه که تعداد نمونه‌های هر کلاس کم می‌باشد، با یادگیری شباهت هر نمونه با الگوهای کلاسی تعمیم‌پذیری مناسبی بر روی داده‌های تست فراهم می‌کند. ProtoNet [۳] ، یک روش طبقه‌بندی اندک-نمونه مبتنی بر الگو است که با استفاده از شبکه‌های عصبی عمیق، بردار ویژگی‌ها را در فضای تعبیه[۱۳] به گونه‌ای محاسبه می‌کند تا میانگین مجموعه نمونه‌های هر کلاس الگوی مثبت مناسبی برای آن کلاس باشد. برچسب نمونه‌ی آزمون، کلاسی می‌شود که نزدیک‌ترین الگوی نماینده به آن را دارد. در [۱۸] برای بهبود روش ProtoNet بیش از یک الگو برای هر کلاس در نظر می‌گیرد. [۱۹] به جای فاصله‌ی اقلیدسی که در کار [۳] استفاده می‌شود، با به‌کار بردن بیز تغییری مترِ فاصله‌ی جدیدی را تعریف می‌نماید.

برخی از مقالات علاوه بر توجه به فاصله تا الگوی نماینده از کلاس صحیح به فاصله‌ای که یک نمونه تا الگوهای نماینده از کلاس-های دیگر دارند نیز توجه داشته‌اند. الگوهای نماینده‌ی کلاس‌های دیگر را به اصطلاح الگوی منفی نام برده‌اند. در ProtoPNet [۵]، الگوهای منفی بکار رفته‌اند که در واقع نزدیک‌ترین الگوی مثبت از کلاس‌های مخالف هستند و با هدف جداپذیری بیشتر کلاس‌ها استفاده می‌شوند. در تحقیقات دیگری نیز در زمینه‌ی یادگیری مبتنی بر الگو که از اصطلاح الگوی منفی استفاده کرده‌اند، منظور از الگوی منفی همان نزدیک‌ترین الگو(ها)ی مثبت از کلاس‌های مخالف است که با نزدیک شدن نمونه‌ها به الگوی مثبتی از کلاس خود و فاصله گرفتن از الگوی منفی، کلاس‌ها تمایزپذیری بیشتری دارند اما مدل طبقه‌بندی همان مدل طبقه‌بندی مبتنی بر الگوی مثبت است [۱۰].

همچنین نمونه یا نماینده‌ی منفی در زمینه‌های دیگر یادگیری از جمله یادگیری مبتنی بر مترِ فاصله[۱۴] برای بالا بردن دقت طبقه‌بندی نزدیک‌ترین همسایه نیز استفاده شده است [۱۲, ۱۱]. استفاده از نمونه‌های منفی در زمینه‌ی یادگیری تضادی نیز رکن اصلی را دارد

[۲۶,۲۷]. در این تحقیقات، هدف یادگیری بدون نظارت بردارهای نماینده‌ی داده‌ها در فضای عمیق به گونه‌ای است که بتوان از این بردارهای نماینده در وظایفی چون طبقه‌بندی استفاده نمود. برای این منظور در این شیوه‌ی یادگیری، بردارهای تعبیه شده به گونه‌ای یادگرفته می‌شود که یک نمونه به نمونه‌ی مثبتی از کلاس خود نزدیک شود و از نمونه یا نمونه‌هایی منفی از کلاس مخالف دور گردد.

±ED-WTA [۲] برای اولین بار با تعریف جدیدی از الگوهای منفی در کنار الگوهای مثبت، مدل جدیدی از طبقه‌بندی مبتنی بر الگو را ارائه داده است که دقت بالاتری نسبت به حالت فقط الگوی مثبت دارد. نکته‌ی قابل توجه، شباهت غیرقابل تشخیص الگوهای مثبت و منفی پس از پایان فاز آموزش است. در پژوهش ارائه شده، با بررسی دلایل این شباهت، به ارتباط مهمی که روش مبتنی بر شبکه عصبی ±ED-WTA و طبقه‌بند SVM دارد، از نقطه‌نظر نمونه‌های مؤثر در تصمیم‌گیری پرداخته می‌شود.

SVM از جمله طبقه‌بندهایی هست که ابرصفحه‌ی جداکننده‌ی کلاس‌ها در آن بر اساس نمونه‌ها شکل می‌گیرد. در طبقه‌بند مبتنی برحاشیه‌ی SVM نمونه‌های محدودی هستند که به عنوان بردارهای پشتیبان انتخاب می‌شوند و ابرصفحه‌ی مبتنی بر حاشیه را می‌سازند [۱۳,۱۶]. همچنین توجه به نمونه‌هایی که در فضای توزیع یک کلاس دور از بقیه نمونه‌ها قرار گرفته‌اند در تحقیقات اخیر در زمینه طبقه-بندی نیز مورد توجه بوده است. مقاله‌ی [۱۷] یک طبقه‌بند تفسیرپذیر مبتنی بر شبکه عصبی عمیق ارائه می‌دهد و برای هر داده‌ی تست، نقاط نماینده مثبت و نقاط نماینده منفی را بر اساس قضیه‌ی نماینده[۱۵] بدست می‌آورد. نقاط نماینده‌ی مثبت برای یک داده‌ی تست یعنی آن نمونه‌هایی از کلاس موافق که علاوه بر شباهت زیادی که دارند، ضریب گرادیان آن‌ها در تعیین پارامتر سطح تصمیم کلاس، مقدار مثبت بزرگی است. نقاط نماینده منفی نمونه‌های آموزشی از کلاس‌های دیگر با شباهت زیاد اما ضرایب گرادیان منفی بزرگ هستند. با توجه به نحوه‌ی محاسبه ضریب تأثیر از روی گرادیان تابع خطای شبکه، داده-های مرزی، ضریب تأثیر بزرگ‌تری دارند. در [۷,۱۵] ، تأثیرگذاری هر نمونه‌ی آموزشی را بر دقتی که شبکه عصبی عمیق برای نمونه‌های آزمون بدست می‌آورد، بررسی کرده است. نویسندگان این مقالات در ادعا می‌کنند در حالتی که توزیع داده ها long-tailed است یعنی نمونه‌های غیرمعمول و پرت به نسبت کم نیستند، به‌خاطرسپاری برچسب برخی داده‌های آموزشی در بالا رفتن دقت طبقه‌بندی نمونه-های تست تأثیر دارد. در نتایج مقاله [۱۵] نشان داده است ضریب اهمیت به خاطر سپاری نمونه‌های کمیاب و پرت بالاتر است.

با توجه به این‌که یافته‌های این مقاله مرتبط با روش ED-WTA± است در بخش ۳ به تفصیل در مورد این روش توضیح داده خواهد شد.

۳- مروری بر روش ED-WTA±

روش ED-WTA± یک مدل طبقه‌بندی تفسیرپذیر مبتنی بر الگو برای شبکه‌های عصبی تک-لایه است. زارعی و همکارانش در [۲] عملکرد نورون مبتنی بر ضرب داخلی را بر اساس اختلاف مربع فاصله‌ی اقلیدسی یک داده با دو الگوی مثبت و منفی تفسیر می‌کنند. در ED-WTA± هر نورون با دو الگو مدل می‌شود: الگوی مثبت که نمایان‌گر نمونه‌هایی است که به درستی با آن نورون درمورد کلاس آن‌ها تصمیم‌گیری می‌شود و الگوی منفی که نمایان‌گر نمونه‌هایی است که به اشتباه، توسط نورون برنده شده‌اند. ساختار شبکه عصبی ED-± WTA در شکل (۱) نشان داده شده است. در این روش با به کاربردن روش [۲۸] برای هر کلاس بیش از یک نورون در لایه‌ی بیشینه‌ی نَرم در نظر گرفته شده است. M تعداد کل نورون‌های در نظر گرفته شده است که برای هر کلاس به طور مساوی تعداد K نورون لحاظ می‌شود. $O_c = \{1, 2, \dots, K\}$ مجموعه‌ی نورون‌های در نظر گرفته شده برای کلاس c می‌باشد.

با توجه به فاصله‌ی یک نمونه تا الگوی مثبت c_j^+ و فاصله‌اش تا الگوی منفی نورون j، c_j^- مقدار فعالیت نورون z_j به صورت رابطه‌ی زیر محاسبه می‌شود:

$$Z_j = -\frac{1}{2}\left(\|x - c_j^+\|^2 - \|x - c_j^-\|^2\right) \quad (۱)$$

غیاثی در [۲۸] به دلیل این‌که برای هر کلاس چند زیرکلاس در نظر می‌گیرد از تابع زیان تعمیم یافته‌ی آنتروپی-متقابل رقابتی[۱۶] (*CCE*) استفاده می‌کند [۲۸]. تابع *CCE* برای انتخاب نورون پیروز یک حالت رقابتی نرم در انتخاب مراکز یک کلاس به‌وجود می‌آورد تا هر نورون مربوط به یک کلاس شانس برنده شدن داشته باشد. تابع *CCE* به صورت زیر تعریف می‌گردد:

$$L_{cce} = -\sum_{n=1}^{N}\sum_{i=1}^{M} \tau_{ni} \log y_{ni} \quad (۲)$$

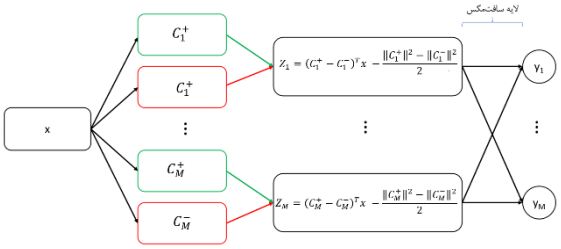
که در این رابطه، τ_{ni} مقداری است که با توجه به توزیع بیشینه-ی نرم حقیقی نمونه‌ها در خوشه‌های یک کلاس تعریف می‌شود:

$$\tau_{ni} = \begin{cases} \dfrac{\exp(\beta z_{ni})}{\sum_l \exp(\beta z_{nl})}, & \forall i, l \in O_c \\ 0, & \forall i \notin O_c \end{cases} \quad (۳)$$

c برچسب صحیح نمونه‌ی n است و y_{ni} در تعریف تابع ضرر CCE خروجی لایه‌ی آخر شبکه عصبی است که احتمال نسبت یک نمونه به مرکز i ام را نشان می‌دهد:

$$y_{ni} = \frac{\exp(\beta z_{ni})}{\sum_{k=1}^{M} \exp(\beta z_{nk})} \quad (۴)$$

در روابط y_{ni} و τ_{ni} ، برای آن که تابع بیشینه‌ی نرم در منطقه‌ی اشباع گرفتار نشود و الگوهای مثبت و منفی به نمونه‌های زیرکلاس خود نزدیک باشند، پارامتر دیگری به نام β نیز در نظر گرفته می‌شود.



شکل (۱): ساختار ED-WTA± [۱]

با محاسبه‌ی گرادیان تابع زیان نسبت به پارامترهای الگوی مثبت و منفی برای هر نورون، بهینه‌سازی الگوها به صورت آنلاین انجام می‌شود. با توجه به برچسب کلاس هر نمونه c، الگوهای مثبت مربوط به نورون‌های کلاس صحیح و الگوهای منفی مربوط به نورون‌های کلاس‌های دیگر به صورت زیر به‌روزرسانی می‌شوند:

$$c_j^{+new} = c_j^{+old} + \mu\beta(\tau_j - y_j)\left(x - c_j^{+old}\right), \forall j \in O_c \quad (۵)$$

$$c_j^{-new} = c_j^{-old} + \mu\beta y_j\left(x - c_j^{-old}\right), \forall j \notin O_c \quad (۶)$$

همچنین پارامتر β نیز با رابطه‌ی زیر بهینه می‌گردد:

$$\beta = \beta - \gamma \sum_{j=1}^{M}(y_j - \tau_j)z_j \quad (۷)$$

در روابط (۵)، (۶) و (۷)، پارامترهای μ و γ نرخ یادگیری هستند.

۴- آنالیز الگوی مثبت و الگوی منفی

در این قسمت با آنالیز روش کاهش گرادیان در مقاله‌ی [۲] و همچنین با ترسیم دقیق نمونه‌های مؤثر در تشکیل الگوهای مثبت و منفی نشان داده می‌شود، تفسیر [۲] در مورد این که الگوی مثبت میانگین نمونه‌های درست دسته‌بندی شده و الگوی منفی میانگین نمونه‌های اشتباه دسته‌بندی شده است، دقیق نمی‌باشد. در مقاله‌ی [۲] دلیل شباهت زیاد الگوهای مثبت و منفی یک کلاس به این صورت آمده است: الگوی مثبت یک کلاس میانگین نمونه‌های منتسب به یک کلاس است و در دسته‌بندی نمونه‌ها، نمونه‌هایی که شباهت زیادی به الگوی مثبت دارند با احتمال بالاتری اشتباه می‌شوند. از برآیند این نمونه‌ها الگوی منفی شکل می‌گیرد که فاصله‌ی کمی با الگوی مثبت دارد. در حالی‌که در پژوهش حاضر نشان داده می‌شود علاوه بر این که الگوی منفی برآیند نمونه‌های منفی شبیه به کلاس موافق است، الگوی مثبت نیز الگوی میانگین نمی‌باشد و برآیند نمونه‌هایی است که شبیه به نمونه‌های کلاس‌های مخالف هستند.

۴-۱- تحلیل آموزش کاهش گرادیانی در ED-WTA±

در این بخش به بررسی و تحلیل نحوه‌ی به‌روزرسانی الگوهای مثبت و منفی نورون‌ها با روش کاهش گرادیان که در روابط (۵) و (۶) آمده است، پرداخته می‌شود. با توجه به تحلیلی که در این رابطه به‌دست آمده است، نام یافته جدید در مورد شباهت تمایز ناپذیر الگوهای مثبت و منفی، کاوش سخت-آگاه الگو نامیده شده است.

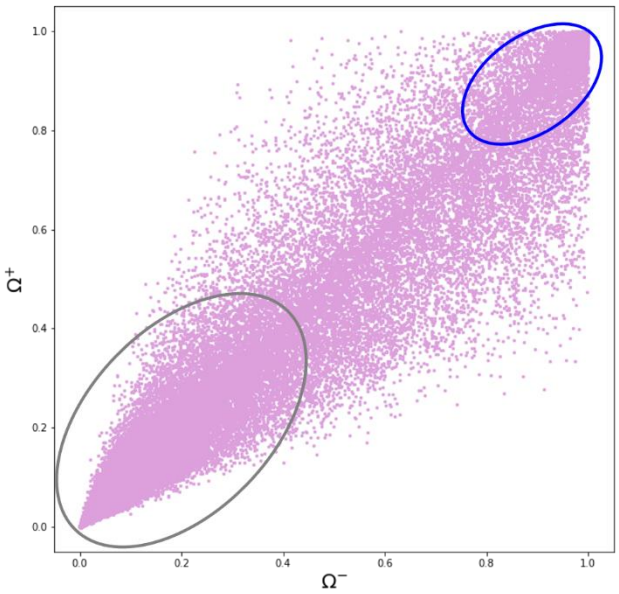
در مورد زیرکلاس j ام با توجه به روابط (۵) و (۶) برای τ_j و y_j، نمونه‌های مثبتی که برای زیرکلاس خود $\tau_j - y_j$ و یا ضریب گرادیان مثبت بزرگی دارند، با توجه به توزیع بیشینه نرم برای y_j، برای زیرکلاس‌های منفی نیز ضریب گرادیان منفی بزرگتری پیدا می‌کنند. به عبارت دیگر نمونه‌های مهم و مؤثر در عین حال که در شکل‌گیری الگوی مثبت کلاس خود تأثیرگذار هستند، این نمونه‌ها در شکل‌گیری الگوهای منفی کلاس‌های دیگر نیز نقش دارند. بزرگترین ضریب گرادیان مثبت یک نمونه Ω^+ و بزرگترین ضریب گرادیان منفی آن، Ω^- به صورت زیر تعریف می‌شود:

$$\Omega^+ = \underset{j\in O_c, 1\le t\le epochs}{\operatorname{argmax}} \tau_{nj}^t - y_{nj}^t \quad (۸)$$

$$\Omega^- = \underset{j\notin O_c, 1\le t\le epochs}{\operatorname{argmax}} y_{nj}^t \quad (۹)$$

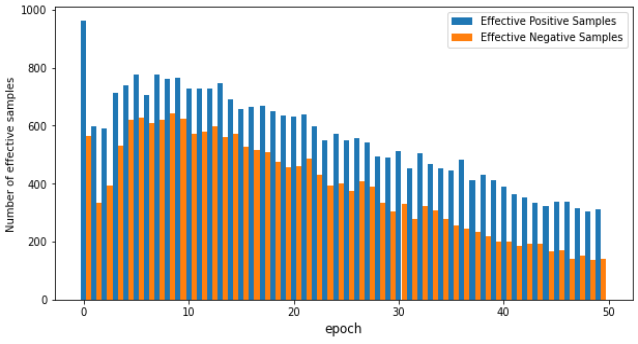
که در روابط (۸) و (۹)، epochs تعداد کل مراحل آموزش مدل است. در نمودار شکل (۲) رابطه‌ی بین Ω^+ و Ω^- برای هر نمونه‌ی MNIST ترسیم شده است. همان‌طور که در شکل مشاهده می‌شود این رابطه برای اکثر نمونه‌ها خطی است. با توجه به این مطلب در ادامه توضیح دقیق‌تری در مورد نحوه‌ی تشکیل الگوهای مثبت و منفی آورده شده است:

- در گام‌های ابتدایی کاهش گرادیان ED-WTA±، نمونه‌هایی تأثیرگذار بر روی الگوها هستند که زیاد از مرکز خوشه‌ها فاصله ندارند داده‌هایی که نزدیک به مرکز خوشه هستند با توجه به رابطه‌ی (۱) امتیاز بیشتری کسب می‌کنند و نمونه‌های مرزی به دلیل این‌که از الگوی مثبت خود هنوز فاصله زیادی دارند امتیاز کمی کسب می‌کنند.
- به تدریج الگوهای مثبت از مرکز خوشه‌ها فاصله می‌گیرند و در این حالت نمونه‌های مرزی و دور از مرکز هستند که امتیاز بیشتری دارند و در تشکیل الگوهای مثبت مؤثر می‌باشند.
- اگر فرض شود الگوهای منفی از حالت اولیه‌ی برابر با میانگین کل داده‌ها شروع می‌شوند، در گام‌های ابتدایی نمونه‌های بیشتری از کلاس‌های دیگر هستند که بر روی الگوی منفی تأثیرگذار هستند اما به تدریج که الگوهای مثبت به الگوهای منفی شبیه‌تر می‌شوند، نمونه‌های منفی محدودی که خیلی شبیه به الگوی مثبت یک کلاس هستند، الگوی منفی آن کلاس را تشکیل می‌دهند.



شکل(۲): رابطه حداکثر گرادیان مثبت (Ω^+) و حداکثر گرادیان منفی (Ω^-) برای هر نقطه داده آموزشی. بسیاری از دادگان نقشی در به‌روزرسانی الگوهای مثبت و منفی ندارند (محصور شده با مدار خاکستری). نمونه‌هایی که نقش مؤثر در شکل‌گیری الگوها دارند، ضریب گرادیان مثبت و منفی بزرگی دارند (محصور شده با مدار آبی)

به عبارت دیگر، به تدریج در گام‌های آموزش از تعداد دادگانی که مؤثر بر الگوهای مثبت و منفی هستند کاسته می‌شود و همین دادگان محدود که تقریباً نزدیک به مرز کلاس‌ها قرار گرفته‌اند هم الگوهای مثبت و هم الگوهای منفی را تشکیل می‌دهند. به همین دلیل نیز الگوهای مثبت و منفی در انتهای آموزش شباهت غیر قابل تشخیصی به یکدیگر دارند. در شکل (۳)، نمودار فراوانی تعداد نمونه‌هایی که ضریب گرادیان مثبت و ضریب گرادیان منفی در بازه‌ی $(0.8, 1)$ داشته‌اند ترسیم شده است. همان‌طور که نمودار نشان می‌دهد در فازهای انتهایی آموزش تعداد نمونه‌های مثبت مؤثر و تعداد نمونه‌های منفی مؤثر نسبت به فازهای ابتدایی کاهش پیدا می‌کند.



شکل(۳): نمودار فراوانی دادگان با بزرگترین ضریب گرادیان مثبت و بزرگترین ضریب گرادیان منفی در بازه $(0.8, 1)$ در ۵۰ گام آموزش

۴- ۲-نمونه‌های با اهمیت

الگوهای مثبت و منفی به صورت میانگین وزن‌داری از نمونه‌ها به صورت روابط (۱۰) و (۱۱) در نظر گرفته می‌شود. با استفاده از مقاله‌ی [1] ضرایب α^+ و α^- یا ضرایب تأثیر نمونه‌ها در بسط الگوها بدست می‌آید:

$$C_j^+ = \sum_{n=1, x_n \in class\ j}^{N} \alpha_{jn}^+ x_n \qquad (۱۰)$$

$$C_j^- = \sum_{n=1, x_n \notin class\ j}^{N} \alpha_{jn}^- x_n \qquad (۱۱)$$

وزن‌های نمونه‌ها در روابط (۱۰) و (۱۱) به صورت برخط با روش کاهش گرادیان به صورت زیر به‌روزرسانی می‌شود:

$$\alpha_j^{+new} = \alpha_j^{+old} + \mu\beta(\tau_j - y_j)\left(\bar{1}_n - \alpha_j^{+old}\right) \qquad (۱۲)$$

$$\alpha_j^{-new} = \alpha_j^{-old} + \mu\beta(y_j)\left(\bar{1}_n - \alpha_j^{-old}\right) \qquad (۱۳)$$

که در این روابط، N تعداد دادگان آموزشی، $\alpha_j^+ \in R^{n\times 1}$ وزن مؤثر نمونه‌ها در تشکیل الگوی مثبت مرکز j ام می‌باشد. $\alpha_j^- \in R^{n\times 1}$ وزن مؤثر نمونه‌ها در بسط الگوی منفی کلاس j ام است. $\bar{1}_n$ برداری است که تنها درایه‌ی n ام آن مقدار ۱ دارد.

نمونه‌های مؤثر مثبت، نمونه‌هایی هستند که در بسط الگوی مثبت یک کلاس در فرمول (۱۰) ضریب α^+ بزرگ و نمونه‌های مؤثر منفی، نمونه‌هایی هستند که در بسط الگوی منفی یک کلاس در فرمول (۱۱) ضریب α^- بزرگی دارند. همچنین گرادیان مثبت یک داده برای مرکز j ام، $\tau_j - y_j$ و گرادیان منفی y_j می‌باشد. با در نظر گرفتن تابع ضرر CCE، رابطه (۴) و همچنین روابط (۵) و (۶)، اگر یک داده آموزشی در نزدیکی مرکز کلاس خود (k^*) قرار داشته باشد، سپس خروجی تابع y_{nk*} تقریباً برابر با مقدار خروجی حقیقی τ_{nk*} است.

به این ترتیب $\tau_{nk*} - y_{nk*}$ (گرادیان مثبت) و همچنین y_{nj} (گرادیان منفی) برای هر مرکز j مربوط به دیگر کلاس‌ها نزدیک به صفر است. اما اگر نمونه‌ی آموزشی دورتر از مرکز کلاس واقعی خود باشد، در این صورت $y_{nk*} \ll \tau_{nk*}$ و در نتیجه $\tau_{nk*} - y_{nk*}$ مقدار بیشتری در هنگام به‌روزرسانی الگوی مثبت کلاس k^* و همچنین y_{nj} مقدار بزرگتری در ارتباط با الگوی منفی کلاس j دارد. به این ترتیب نمونه‌های مرزی گرادیان مثبت و منفی بزرگتری دارند یعنی نقش بیشتری در به‌روزرسانی الگوها دارند. همچنین با توجه به روابط (۱۲) و (۱۳) ضریب تأثیر بالاتری نیز در بسط الگوهای مثبت و منفی دارا هستند.

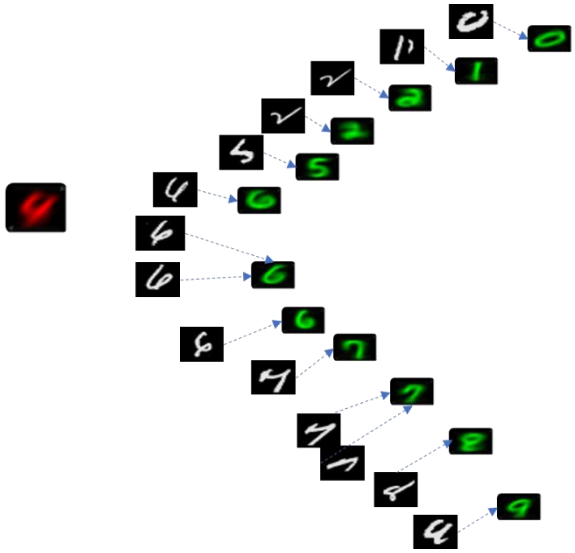
یک نمونه ممکن است ضریب تأثیر بالایی در تشکیل یک الگوی مثبت داشته باشد و همین نمونه در تشکیل الگوی منفی از یک کلاس دیگر نیز مؤثر باشد. برای نشان دادن این موضوع در شکل (۴)، نمونه‌های ارقام 4 از مجموعه دادگان MNIST نشان داده شده است که

ضریب اهمیت بالا(α^+) در تشکیل الگوی مثبت خوشه دوم این کلاس دارند.

شکل (۴): تصویر ۳۰ نمونه مثبت با بالاترین ضریب تأثیر برای الگوی مثبت خوشه دوم از کلاس ۴ (مرکز شماره ۲۵)

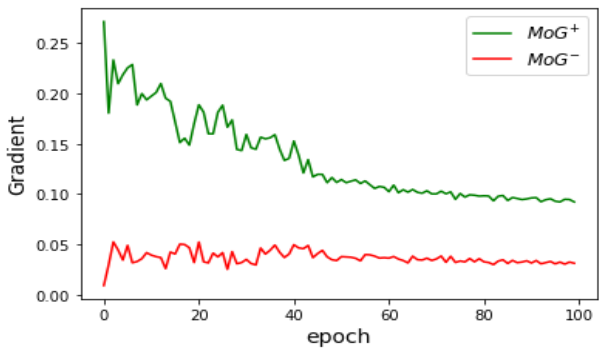
در شکل (۵) تعدادی از این نمونه‌های مثبت، که ضریب تأثیر بالایی در تشکیل الگوی منفی مراکز کلاس‌های دیگر دارند نیز نشان داده شده است. تصویر سبز، الگوی مثبت از کلاس 4 را نشان می‌دهد. این الگو با توجه به نمونه‌هایی از کلاس 4 ساخته می‌شود که نمونه‌های دقیقی از این کلاس نیستند یا به عبارت دیگر در توزیع کلاس، نزدیک مرکز قرار نمی‌گیرند. از طرف دیگر این نمونه‌ها در تشکیل الگوهای منفی کلاس‌های مخالف که با رنگ قرمز نمایش داده شده‌اند نیز تأثیرگذار می‌باشند.

شکل (۵): برخی از نمونه‌های مثبت خوشه دوم کلاس رقم 4 که در شکل‌گیری الگوهای منفی (تصاویر قرمز) از کلاس‌های دیگر نیز با ضریب تأثیر بالا نقش داشته‌اند.



شکل (۶): نمونه‌های منفی با ضریب تأثیر بالا در شکل‌گیری الگوی منفی خوشه دوم از کلاس 4 (شکل 4 به رنگ قرمز) که نمونه‌های مثبت مؤثر برای الگوهای مثبت مراکز دیگر (ارقام سبز رنگ) هستند.

در شکل (۶) مشاهده می‌شود، الگوی منفی کلاس 4 نیز با توجه به نمونه‌هایی ساخته شده است که در شکل‌گیری الگوهای مثبت کلاس‌های دیگر ضریب تأثیر بالایی داشته‌اند. نکته‌ی قابل توجه دیگر اشتراک اکثریت الگوهای منفی (ارقام به رنگ قرمز) در شکل (۵) و الگوهای مثبت (ارقام به رنگ سبز) در شکل (۶) است. به عبارت دیگر این اشتراک، تشابه نمونه‌های مربوط به الگوهای مثبت و منفی برای هر مرکز را نیز نشان می‌دهد که باعث تمایز ناپذیر شدن الگوهای مثبت و منفی می‌شود.

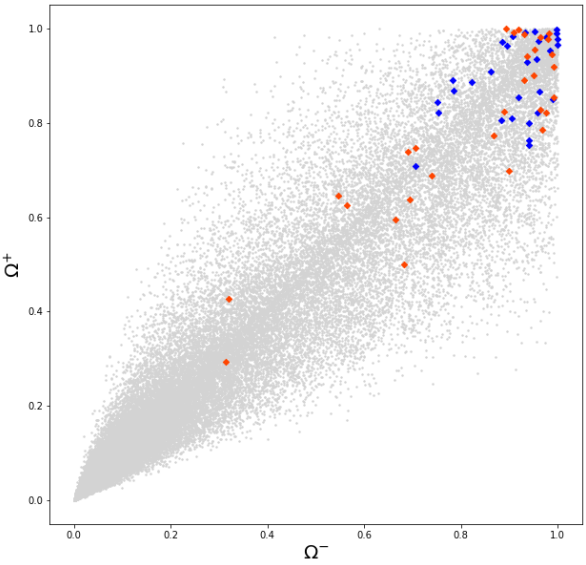


شکل(۷): میانگین گرادیان نمونه‌های مثبت مؤثر متعلق به الگویی از کلاس 4(MoG^+) و گرادیان منفی همین نمونه‌ها به ازای الگویی از کلاس 8 (MoG^-)

در شکل (۵)، نمونه‌های مثبت از کلاس 4 (خوشه ۲۵ ام) با الگوی منفی از کلاس 8 (خوشه ۵۳) اشتراک بیشتری دارند. در نمودار شکل (۷) میانگین ضرایب گرادیان مثبت این نمونه‌ها به ازای الگوی مثبت کلاس 4 (MoG^+) و گرادیان منفی (MoG^-) آن‌ها به ازای الگوی منفی از کلاس 8 رسم شده است. در این شکل ملاحظه می‌شود

در مراحل انتهایی آموزش تغییرات میانگین گرادیان مثبت و منفی این نمونه‌ها به یکدیگر نزدیک‌تر می‌شود.

نمونه‌هایی که در بسط الگوی مثبت و بسط الگوی منفی خوشه دوم از کلاس 4، ضریب α بزرگی دارند، در شکل(۸) نمایش داده شده است. برای نمونه‌های مثبت (نقاط به رنگ آبی) همان‌طور که در شکل مشخص است، ضریب گرادیان Ω^+ و ضریب گرادیان Ω^- نزدیک به ۱ است.



شکل (۸): نمایش نمونه‌های مثبت با ضریب تأثیر بالا در بسط الگوی مثبت (نقاط آبی) و نمونه‌های منفی مؤثر در تشکیل الگوی منفی (نقاط قرمز) مربوط به خوشه دوم از کلاس۴ بر روی نمودار نسبت گرادیانی

با تحلیل گرادیانی که در این بخش ارائه شده است، می‌توان به این نتیجه رسید که برای بدست آوردن مدل یادگیری مبتنی بر الگوهای مثبت و منفی، در پایان آموزش نمونه‌های محدودی نیاز هست تا با آن‌ها الگوی مثبت و الگوی منفی را برای هر مرکز بدست آورد. نمونه‌هایی که در فازهای آخر بر روی الگوها تأثیرگذارتر هستند، نمونه‌های مرزی هستند که هم ضریب گرادیان مثبت و هم ضریب گرادیان منفی بزرگی دارند.

۴- ۳- ارتباط روش ED-WTA± و SVM

SVM ابرصفحه‌ی جداکننده‌ی دوکلاس با بیشترین حاشیه را پیدا می‌کند. در این طبقه‌بند غیر از نقاطی که درست دسته‌بندی نشده‌اند، نقاطی نیز جریمه می‌شوند که با حاشیه اطمینان کمتر از ۱ در کلاس درست خود قرار گرفته‌اند. مسئله‌ی بهینه‌سازی SVM دو کلاسه، به صورت رابطه‌ی لاگرانژ زیر تعریف می‌شود:

$$L(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$-\sum_{i=1}^{N} A_i\left(y_i(W^T x_i + b - 1 + \xi_i)\right) - \sum_{i=1}^{N}\eta_i\xi_i \qquad (۱۴)$$

که در رابطه‌ی بالا ترم اول، نزدیک‌ترین فاصله‌ی نقاط هر دو کلاس از ابرصفحه جداکننده که با $\frac{1}{\|w\|}$ تعیین می‌شود را ماکزیمم و یا به طور معادل $\frac{1}{2}\|w\|^2$ را مینیمم می‌کند. قید مربوط به رعایت حاشیه‌ی اطمینان با در نظر گرفتن پارامتر جریمه ξ_i نیز به رابطه اضافه می‌گردد. همچنین باید مجموع جرایم کمینه شود و نیز $\xi_i \geq 0$.

با محاسبه گرادیان تابع لاگرانژ نسبت به پارامترهای (w, ξ_i, b) نتایج مهم زیر حاصل می‌شود. نکات مهمی در این قسمت وجود دارد که در بحث و آنالیزهای بعدی مورد استفاده قرار خواهد گرفت. همچنین با توجه به این نتایج، دوگان مسئله SVM بدست می‌آید که با حل آن، ضرایب بردارهای پشتیبان یعنی A_i ها حاصل می‌شود. در نوشتن روابط (۱۴)، (۱۵) و (۱۶) از مقاله‌ی [۱۶] استفاده شده است.

محاسبه مشتق تابع لاگرانژ نسبت به پارامترهای (w, ξ_i, b) :

$$w = \sum_{i=1}^{N} y_i A_i x_i = \sum_{i\in I^+} A_i x_i - \sum_{i\in I^-} A_i x_i \qquad (۱۵)$$

$$\sum_{i=1}^{N} A_i y_i = 0 \leftrightarrow \sum_{i\in I^+} A_i = \sum_{i\in I^-} A_i \qquad (۱۶)$$

که می‌توان بدون تغییر در جواب مسئله که جهت w می باشد، این فرض را در نظر گرفت:

$$\sum_{i\in I^+} A_i = \sum_{i\in I^-} A_i = 1$$

یعنی مجموع ضرایب بردارهای پشتیبان برای نمونه‌های مثبت از یک کلاس و نمونه‌های منفی از کلاس دیگر برابر با یک می‌باشد. به عبارت دیگر، ضرایب بردار پشتیبان یک مرکز بر روی مجموعه‌ی نمونه‌ها توزیع نرمال دارد. این موضوع برای حالت SVM چندکلاسه نیز در مقاله‌ی [۲۰] اثبات شده است.

در روش ED-WTA± ، نورون مدل شده توسط الگوی مثبت و الگوی منفی مانند یک ابرصفحه‌ی دوکلاسه عمل می‌کند به طوری‌که نمونه‌های کلاس مثبت در سمت الگوی مثبت قرار می‌گیرند و نمونه‌های دیگر کلاس‌ها در سمت الگوی منفی هستند. حال برای تشریح این مطلب که ابرصفحه‌ی جداساز توسط کدام نمونه‌ها ساخته می‌شود، روابط این روش، به شیوه‌ای تحلیل می‌گردد تا ارتباط آن با روش SVM نیز آشکار شود. امتیاز هر نورون در روش ED-WTA± به صورت زیر بازنویسی می‌گردد:

$$Z_j = \left(c_j^+ - c_j^-\right)^T x_n - \left(\frac{\left\|c_j^+\right\|^2 - \left\|c_j^-\right\|^2}{2}\right) \qquad (۱۷)$$

ضریب نگاشت هر داده $\left(c_j^+ - c_j^-\right)$ می‌باشد که با لحاظ کردن روابط (۱۰) و (۱۱) به صورت زیر بدست می‌آید:

$$w_j = \left(c_j^+ - c_j^-\right) = \sum_{i=1}^{N}\left(\alpha_{ji}^+ - \alpha_{ji}^-\right) x_i \qquad (۱۸)$$

با توجه به روابط (۱۲) و (۱۳) مربوط به به‌روزرسانی ضریب نمونه‌ها اثبات می‌شود در روش ED-WTA± نیز مجموع ضرایب مثبت و منفی برابر یک است. یعنی:

$$\sum_{n\in class_j}\alpha_{nj}^+ = \sum_{n\in class_j}\alpha_{nj}^- = 1 \qquad (۱۹)$$

اثبات این مطلب به این ترتیب انجام می‌شود؛ در مقداردهی اولیه، الگوی مثبت، مرکز خوشه است و الگوی منفی میانگین تمام داده‌ها، پس مقدار اولیه ضرایب به ترتیبی است که مجموع آنها ۱ است. با توجه به روابط به‌روزرسانی وزن‌های داده‌ها در روابط (۱۲) و (۱۳) اثبات می‌گردد مقداری که به مجموع ضرایب مثبت یا مجموع ضرایب منفی در هر مرحله اضافه می‌شود صفر است در نتیجه همواره رابطه‌ی (۲۱) برقرار است. برای ضرایب نمونه‌ها در بسط الگوی مثبت، رابطه‌ی زیر حاصل می‌شود:

$$\mu\beta\left(\tau_j - y_j\right)\left(-\alpha_{jn}\right) + \sum_{i\neq n}\mu\beta\left(\tau_j - y_j\right)\left(-\alpha_{ji}\right)$$
$$= \mu\beta\left(\tau_j - y_j\right)\left(\sum_{i=1}^{N} -\alpha_{ji}\right) + \mu\beta\left(\tau_j - y_j\right) = 0$$

لازم به ذکر است که در عبارت سمت راست تساوی:

$$\sum_{i=1}^{N} -\alpha_{ji} = -1$$

به این ترتیب مجموع ضرایب مثبت همواره ۱ است و می‌توان همین اثبات را در مورد مجموع ضرایب در بسط الگوی منفی نیز نشان داد. با توجه به روابط (۱۷)، (۱۸) و (۱۹) می‌توان گفت با تعریف الگوی مثبت و الگوی منفی برای هر زیرکلاس، مرز تصمیم‌گیری توسط این الگوها مانند یک SVM دو کلاسه تعریف می‌شود.

نسخه‌ی SVM چند کلاسه و چند الگویه در مقاله‌ی [۲۲] ارائه شده است، که در این روش تعداد قیدهای مسئله زیاد و یک مسئله بهینه‌سازی غیرمحدب است. در مقاله‌ی یادگیری برخط [۲۱] عنوان طبقه‌بند چند کلاسه و چند الگویه آمده است که البته منظور از الگو با مطلبی که در این پژوهش آمده است متفاوت می‌باشد. در واقع SVM چند کلاسه و چند برچسبی می‌باشد و برای هر کلاس تنها یک الگو درنظر می‌گیرد. [۲۱] طبقه‌بند مبتنی بر حاشیه‌ی چند کلاسه را به صورت چند طبقه‌بند دوکلاسه مدل کرده است. کلاس مورد نظر، کلاس مثبت و دیگر کلاس‌ها، کلاس منفی هستند. در این مقاله از نسخه SVM چند الگویه [۲۵] که به شیوه‌ی برخط وزن‌های نورون‌های برنده برای کلاس موافق و مخالف را به‌روزرسانی می‌نماید، به منظور شبیه‌سازی آزمایشات استفاده می‌شود. روابط در خصوص به‌روزرسانی وزن نورون‌ها به صورت زیر می‌باشد:

$$W_p^{new} = W_p^{old} + \lambda x_n\,,\quad p = \underset{j\in O_{class\ y_n}}{\operatorname{argmax}}\ W_j^T x_n \qquad (۲۰)$$

$$W_q^{new} = W_q^{old} - \lambda x_n\,,\quad q = \underset{j\notin O_{class\ y_n}}{\operatorname{argmax}}\ W_j^T x_n \qquad (۲۱)$$

در روابط (۲۰) و (۲۱)، W_p نورون برنده از کلاس موافق و W_q نورون برنده از کلاس مخالف می‌باشد. λ نیز نرخ یادگیری می‌باشد.

البته برخی از تحقیقات در زمینه ماشین بردار پشتیبان هستند که به نمونه‌هایی که دورتر از مرز قرار گرفته‌اند توجه خود را معطوف کرده‌اند. در [۲۳] مقداری که برای حاشیه اطمینان در نظر گرفته می‌شود با پارامتر ν تنظیم می‌گردد. که ν حدپایینی برای تعداد بردارهای پشتیبان و همچنین حدبالایی برای تعداد بردارهای خطا است. به این ترتیب بهینه‌سازی مقدار حاشیه با توجه به پارامتر ν این امکان را می‌دهد که بردارهای پشتیبان از مرز کلاس‌ها فاصله‌ی بیشتری داشته باشند. در مقاله‌ی [۲۴] نیز که نسخه خلوت طبقه‌بند SVM را ارائه می‌دهد، بردارهای پشتیبان در فاصله‌ای دورتر از مرز کلاس‌ها و نزدیک به مرکز کلاس‌ها قرار دارند.

استراتژی انتخاب نمونه‌های مؤثر در روش ED-WTA± مشابه استراتژی انتخاب بردارهای پشتیبان در روش SVM بر مبنای انتخاب نمونه‌های مرزی است که البته روش مبتنی بر الگوهای مثبت-منفی در انتهای آموزش تعداد نمونه‌های مؤثر کمتری را برمی‌گزینند. این مطلب در شکل (۳) بخش ۴-۱ نشان داده شده است.

۵- آزمایشات

در این بخش نتایج یافته‌های پژوهش حاضر بر روی مجموعه داده‌های واقعی MNIST ، FERET و Fashion-MNIST آمده است.

MNIST یک مجموعه داده‌ی استاندارد برای ارزیابی مسائل طبقه‌بندی چندکلاسه است. این مجموعه داده شامل ۶۰۰۰۰ تصویر آموزشی و ۱۰۰۰۰ تصویر آزمون است. این تصاویر مربوط به ۱۰ رقم دست‌نویس انگلیسی است که سایز ۲۸×۲۸ پیکسل دارند.

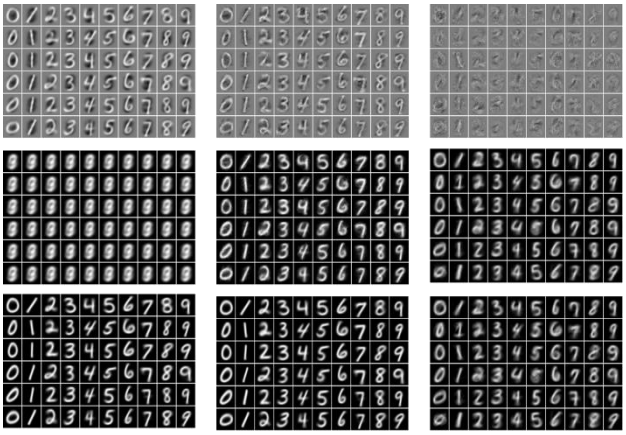
مجموعه داده‌ی FERET یک مجموعه داده‌ی دو کلاسه شامل ۱۴۷۶ تصویر چهره می‌باشد. ۸۲۲ چهره مربوط به مردان و ۶۵۴ چهره مربوط به زنان است. تصاویر پس از ویرایش ۱۹۲×۱۹۲ پیکسل هستند.

مجموعه داده‌ی دیگری که در آزمایشات مورد ارزیابی قرار می‌گیرد، مجموعه داده‌ی Fashion-MNIST(Fashion) می‌باشد. مجموعه داده‌ی آموزشی Fashion شامل ۶۰۰۰۰ تصویر البسه و مجموعه داده‌ی آزمون آن، شامل ۱۰۰۰۰ تصویر است. به هر داده برچسبی نسبت داده شده است که تعلق آن را به یکی از ۱۰ کلاس البسه مشخص می‌کند. تصاویر Fashion همانند تصاویر مجموعه داده‌ی MNIST، ۲۸×۲۸ پیکسل می‌باشند.

در آزمایشاتی که انجام شده است، با توجه به گزارشات مقاله‌ی [۲]، برای هر کلاس از مجموعه دادگان MNIST و همچنین Fashion تعداد k=6 خوشه و برای هر کلاس از مجموعه داده‌ی

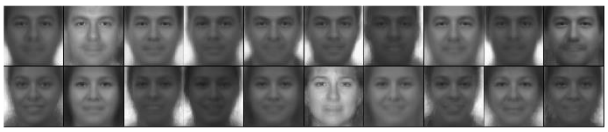
FERET، k=10 خوشه در نظر گرفته شده است. مراکز خوشه‌ها با الگوریتم خوشه بندی kmeans تعیین می‌شود و به عنوان مقدار اولیه-ی الگوهای مثبت در نظر گرفته می‌شود. میانگین کل دادگان، حالت اولیه‌ی الگوهای منفی را تعیین می‌کند. همچنین نرخ یادگیری μ=0.01، γ=0.001، مقدار اولیه پارامتر $\beta = 1$ و تعداد مراحل آموزش، epochs=50 است.

برای بررسی موضوع شباهت الگوهای مثبت و منفی با توجه به توضیحات مقاله‌ی [۲] روشی غیرآنلاین به نام Mean-PNPL[۱۷] پیاده-سازی شده است که در هر گام آموزش، الگوی مثبت میانگین نمونه-های درست دسته‌بندی شده و الگوی منفی میانگین نمونه‌هایی که با یک نورون خروجی به اشتباه دسته‌بندی می‌شوند، می‌باشد. در واقع در این روش تفسیری که [۲] در مورد الگوی مثبت و الگوی منفی و دلیل شباهت زیاد آن‌ها ارائه داده است، لحاظ شده است. در نتایج مشاهده می‌شود که الگوهای مثبت از میانگین نمونه‌های زیرکلاس فاصله‌ی ناچیزی می‌گیرند و تنها الگوهای منفی تا حدی از حالت اولیه که میانگین کل دادگان هست فاصله گرفته و به مرکز زیرکلاس نزدیک می‌شوند. در روش Mean-PNPL دقت نسبت به حالتی که آموزش فقط با الگوی مثبت باشد بهبود کمی پیدا می‌کند. در شکل (۹) مشاهده می‌شود، فاصله‌ی الگوهای مثبت و منفی در این حالت نسبت به نتایج ED-WTA± به طور واضح بیشتر است.



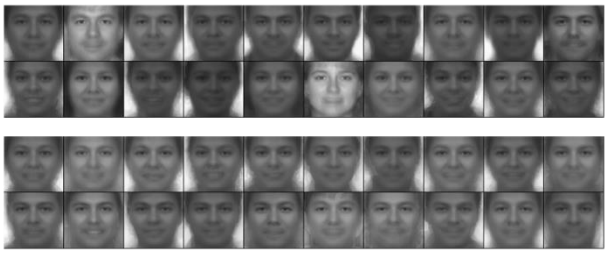
شکل (۹): نتایج روش Mean-PNPL و ED-WTA± بعد از ۱۰۰ مرحله آموزش- ستون سمت چپ مقدار اولیه الگوهای مثبت و منفی- ستون وسط شکل الگوهای حاصل از روش Mean-PNPL و ستون سمت راست الگوهای حاصل از روش ED-WTA± است. تصاویر اولین ردیف اختلاف الگوهای مثبت و منفی متناظر رانشان می‌دهد، تصاویر وسط الگوهای منفی و تصاویر آخرین ردیف الگوهای مثبت برای مراکز MNIIST را نشان می‌دهد.

برای مجموعه دادگان FERET که شامل دو کلاس زنان و مردان است، ۱۰ مرکز برای هر کلاس در نظر گرفته شده است. مراکز با روش Kmeans مقداردهی اولیه شده‌اند که در شکل (۱۰) نمایش داده شده است.

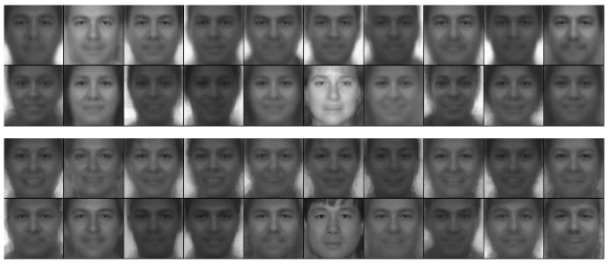


شکل(۱۰): حالت اولیه‌ی ۲۰ مرکز مجموعه دادگان FERET

نتایج روش ED-WTA± مربوط به مجموعه داده‌ی FERET در شکل (۱۱) نمایش داده شده است. در این شکل دو ردیف بالا الگوهای مثبت مربوط به ۲۰ مرکز FERET را نشان می‌دهد و در دو ردیف پایین الگوهای منفی برای این مراکز ترسیم شده است. در شکل (۱۲) نیز حاصل روش Mean-PNPL بر روی مجموعه داده FERET ترسیم شده است.



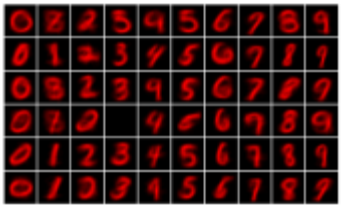
شکل(۱۱): الگوهای مثبت (ردیف بالا) و الگوهای منفی (ردیف پایین) حاصل از روش ED-WTA± مربوط به مجموعه داده FERET



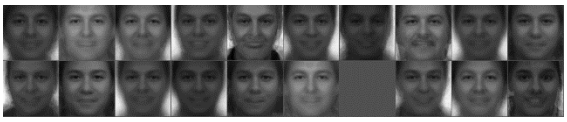
شکل(۱۲): الگوهای مثبت (ردیف بالا) و الگوهای منفی (ردیف پایین) حاصل از روش Mean-PNPL مربوط به مجموعه داده FERET

۵-۱- مقایسه‌ی نمونه‌های مثبت و منفی با بردارهای پشتیبان SVM چندالگویه

در این قسمت، نتایج حاصل از طبقه بند SVM چند الگویه-چندکلاسه که با روش آنلاین آموزش داده شده است برای مقایسه با روش ED-WTA± بررسی می‌شود. در این حالت نیز برای هر کلاس تعداد k=6 خوشه در نظر می‌شود. حالت اولیه برای مراکز نیز با خوشه‌بندی Kmeans تعیین می‌شود. در شکل (۱۳) میانگین بردارهای پشتیبان با ضرایب غیرصفر برای هر مرکز مجموعه داده MNIST و در قسمت (ب) برای مجموعه داده‌ی FERET رسم شده است.



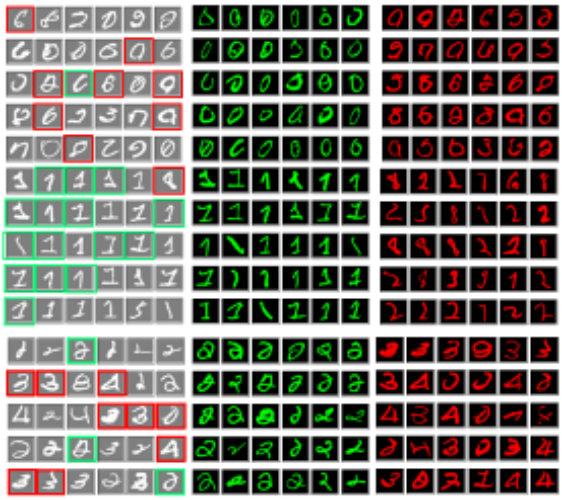
الف



ب

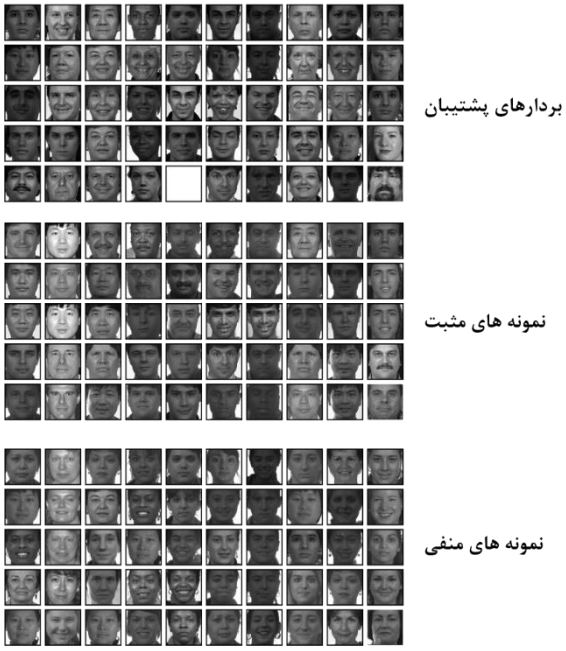
شکل(۱۳): میانگین بردارهای پشتیبان به ازای هر مرکز الف) دادگان MNIST ب)دادگان FERET

همان طور که در این شکل مشاهده می‌شود، برخی مراکز به دلیل این که یا هیچ‌گاه برنده نمی‌شوند و یا با حاشیه اطمینان مناسب نمونه‌ها را طبقه بندی می‌کنند، بردار پشتیبانی ندارند و میانگین صفر دارند. برای مقایسه‌ی عملکرد SVM چندالگویه در انتخاب بردارهای پشتیبان با عملکرد روش ED-WTA±، برای مراکزی از کلاس ۰، ۱ و ۲، در شکل (۱۴) بردارهای پشتیبان و نمونه‌های مثبت-منفی مؤثر به ترتیب ضریب اهمیت رسم شده است. در تصویر بردارهای پشتیبان مشترک با نمونه‌های مثبت مؤثر و نمونه‌های منفی مؤثر به ترتیب با کادر سبز و قرمزرنگ مشخص شده‌اند.



شکل(۱۴): ستون سمت چپ- بردارهای پشتیبان با بزرگ‌ترین ضریب برای ۳ زیرکلاس از کلاس های 0، 1، 2. ستون وسط و سمت راست: به ترتیب ۳۰ نمونه‌ی مثبت و ۳۰ نمونه‌ی منفی با بزرگ‌ترین ضریب در بسط الگوهای مثبت و منفی زیرکلاس‌های متناظر است

مقایسه‌ی بردارهای پشتیبان SVM چند الگویه و نمونه‌های با اهمیت در بسط الگوی مثبت و منفی برای مجموعه داده FERET و کلاس مردان در شکل (۱۵) نشان داده شده است. در این شکل در هر مورد ۵ نمونه از هر مرکز انتخاب شده است. همچنین برای کلاس زنان نیز در شکل (۱۶) این مقایسه انجام شده است. به دلیل این‌که SVM چند الگویه تنها مراکزی که حاشیه اطمینان را رعایت نکرده‌اند جریمه می‌کند، در شکل (۱۵) و (۱۶) مشاهده می‌شود برخی مراکز تعداد بردارهای پشتیبان کمتری دارند و یا بردار پشتیبان ندارند.



شکل(۱۵): پنج نمونه با بالاترین ضریب تأثیر انتخاب شده به عنوان بردارهای پشتیبان در روش SVM. مؤثرترین نمونه‌های مثبت و نمونه‌های منفی در روش ED-WTA± برای ۱۰ مرکز کلاس مردان. هر ستون نمایش دهنده‌ی یک مرکز است

در جدول (۱) نتایج آزمایش روش ED-WTA± در حالتی که فقط الگوهای مثبت آموزش داده شوند و یا تنها الگوهای منفی در مدل قرار داشته باشند مورد بررسی قرار گرفته است. همچنین نتیجه‌ی آزمایش روشی که تنها یک الگو برای هر مرکز لحاظ می‌شود و تمام نمونه‌ها بدون درنظر گرفتن برچسب کلاسی در به‌روزسانی آن تأثیر می‌گذارند نیز لحاظ شده است. نتایج دقت حاصل از این آزمایشات به ویژه بر روی دادگان MNIST و FERET نشان می‌دهد در صورتی‌که برای هر مرکز تنها الگوی منفی درنظر گرفته شود که نمونه‌های کلاس‌های دیگر (نمونه‌های منفی) در تشکیل آن نقش داشته باشند دقت بهتری حاصل می‌شود. در حالی‌که با درنظر گرفتن فقط الگوهای مثبت حاصل از برآیند نمونه‌های مثبت برای آن مرکز، دقت پایین‌تری به‌دست می‌آید. در آموزش تنها یک الگو برای هر مرکز که با توجه به تمام نمونه‌ها به‌روزرسانی می‌شود دقتی بالاتر از آموزش تنها الگوی مثبت و یا تنها الگوی منفی کسب می‌شود. اما الگوی حاصل دیگر شبیه به نمونه‌های کلاسی نمی‌باشد.

از طرف دیگر در جدول (۱)، دقت روش ED-WTA± که در [۲] گزارش شده است با دقت حاصل از روش Mean-PNPL مقایسه شده است. در روش Mean-PNPL الگوهای مثبت و منفی بر اساس تفسیر

نادقیقی که در [۲] بیان شده است، آموزش داده می‌شوند. با مقایسه‌ی دقت حاصل از این روش با ±ED-WTA مشاهده می‌شود در حالت Mean-PNPL که الگوهای مثبت یک نورون میانگین نمونه‌های درست دسته‌بندی شده توسط آن نورون باشند و الگوهای منفی نیز میانگین نمونه‌های به اشتباه دسته‌بندی شده باشند، دقت پایین‌تری حاصل می‌شود. همچنین در نتایج دقت حاصل از این روش در جدول (۱) مشاهده می‌شود، آموزش الگوهای منفی با نمونه‌های به اشتباه دسته‌بندی شده در بالا بردن دقت، تأثیرگذاری بیشتری دارد. روش SVM چندالگویه نیز بر روی مجموعه دادگان مورد بررسی، پیاده‌سازی شده است که دقت حاصل از این روش در جدول (۱) گزارش شده است. در شرایطی که مراکز خوشه‌ها از مقدار اولیه‌ی مناسبی شروع شوند دقت به طور قابل ملاحظه نسبت به حالت اولیه‌ی تصادفی بهتر است.

نتایج نشان می‌دهد ±ED-WTA با تابع زیان CCE دقتی در حدود ۹۶,۴۴٪ بر روی مجموعه داده MNIST، دقت ۹۷,۲۹٪ بر روی مجموعه داده FERET و دقتی در حدود ۸۷,۹۴٪ بر روی مجموعه دادگان Fashion به‌دست می‌آورد.

۶- نتیجه‌گیری

روش ±ED-WTA بر اساس اختلاف مربع فاصله‌ی اقلیدسی از الگوی مثبت و الگوی منفی عملکرد یک نورون را تفسیر می‌کند. در این مقاله با تحلیل روش طبقه‌بندی مبتنی بر الگوی ±ED-WTA تعریف جدیدی از الگوی مثبت و الگوی منفی بر اساس نمونه‌های مؤثر در شکل‌گیری آن‌ها ارائه شد. در یافته‌ی جدید، نشان داده شد با توجه به این‌که الگوهای مثبت و منفی از روی نمونه‌هایی ساخته می‌شوند که نزدیک مرز کلاس‌ها قرار گرفته‌اند، شباهت بسیار زیادی با یکدیگر دارند. همچنین نظریه‌ی جدید بیان کرد، به طور ضمنی نورون‌ها در لایه‌ی رقابتی شبکه عصبی مشابه SVM بر اساس داده‌های مرزی تصمیم‌گیری را انجام می‌دهند.

نظریه‌های ارائه شده در این مقاله به آن دسته از تحقیقاتی نظیر یادگیری اندک-نمونه[۱۸] و یادگیری مترِ فاصله که بر اساس تابع طبقه‌بندی بیشینه‌ی نرم مبتنی بر فاصله‌ی اقلیدسی هستند، کمک می‌کند تا با به‌کاربردن الگوهای منفی در کنار الگوهای مثبت دقت بالاتری را کسب نمایند.

شکل(۱۶): پنج نمونه با بالاترین ضریب تأثیر انتخاب شده به عنوان بردارهای پشتیبان، نمونه‌های مثبت، نمونه‌های منفی برای ۱۰ مرکز کلاس زنان- هر ستون نمایش دهنده‌ی یک مرکز است

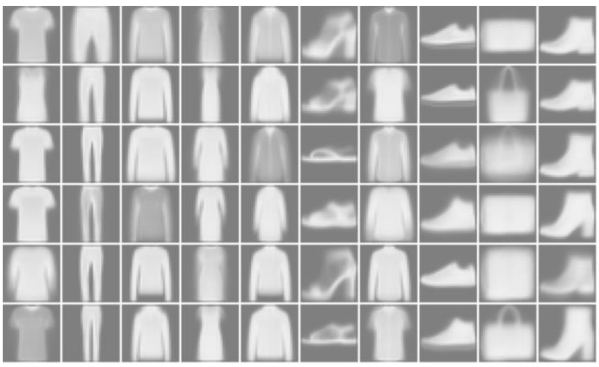
جدول(۱): نتایج دقت حاصل از روش ±ED-WTA، Mean-PNPL و SVM چند الگویه

روش	نحوه‌ی آموزش الگوها	MNIST	FERET	Fashion
±ED-WTA	حالت ورودی KMeans	٪۹۱,۰	٪۸۳,۷۸	٪۷۷,۲۱
	آموزش فقط الگوهای مثبت با نمونه‌های مثبت	٪۹۲,۸۴	٪۸۱,۰۸	٪۷۹,۱۴
	آموزش تنها یک الگو برای هر مرکز با تمام نمونه‌ها	۹۵,۷۵	٪۸۲,۴۳	٪۸۵,۹۲
	آموزش فقط الگوهای منفی با نمونه‌های منفی	٪۹۲,۸۷	٪۸۸,۵۱	٪۷۹,۱۵
	آموزش الگوهای مثبت و منفی	**٪۹۶,۴۴**	**٪۹۷,۲۹**	**٪۸۷,۹۴**
Mean-PNPL	آموزش فقط الگوهای مثبت-میانگین	٪۹۰,۴	٪۷۹,۷۳	٪۷۶,۵
	آموزش فقط الگوهای منفی-میانگین	۹۳,۰۳	٪۸۰,۴	۷۹,۲۰
	آموزش الگوهای مثبت-میانگین و منفی-میانگین	٪۹۲,۸۳	٪۹۰,۵۸	٪۸۰,۳
MultiPrototype-SVM	حالت اولیه‌ی تصادفی	٪۹۴,۹۶	٪۹۱,۸۹	٪۸۳,۷۴
	حالت اولیه‌ی KMeans	٪۹۵,۲۸	٪۹۳,۹۲	٪۸۲,۵۸

ضمایم

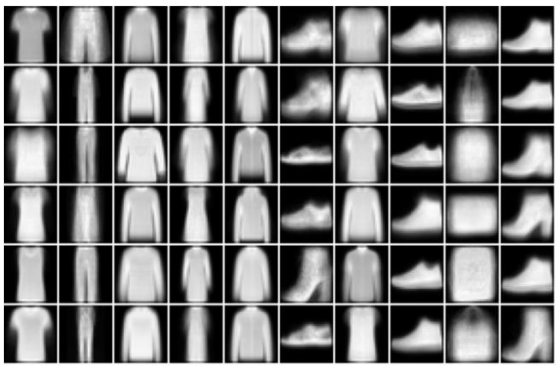
در ایـن بخـش نتـایج آزمـایش روش‌هـای ED-WTA±، SVM چندالگویـه و Mean-PNPL بـر روی مجموعـه دادگـان Fashion-MNIST ارائه شده است.

برای پیاده‌سازی روش ED-WTA± برای هر کلاس ۶ زیـرکلاس در نظر گرفته شده است و مقدار اولیه‌ی مراکز بـا الگـوریتم Kmeans تعیین می‌شود. شکل (۱۷) مقدار اولیه‌ی مراکز برای این مجموعه داده را نشان می‌دهد.

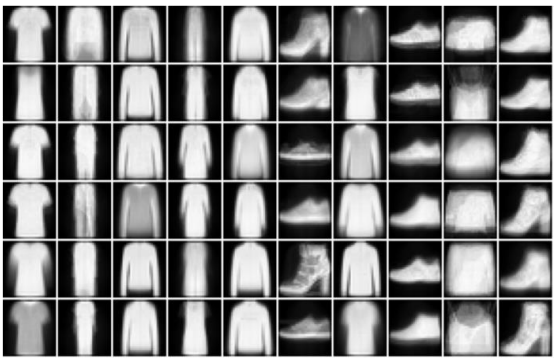


شکل (۱۷): حالت اولیه‌ی مراکز مجموعه دادگان FashionMNIST

در شکل (۱۸) و (۱۹) الگوهای مثبت و منفی منتج شده از روش ED-WTA± بر روی این مجموعه داده رسم شده است. بـر روی ایـن مجموعه داده نیز بیشترین تفاوتی که الگوی مثبت و منفی یـک مرکـز دارند و به سختی قابل تشخیص است، محوتر بودن الگوی منفی است.

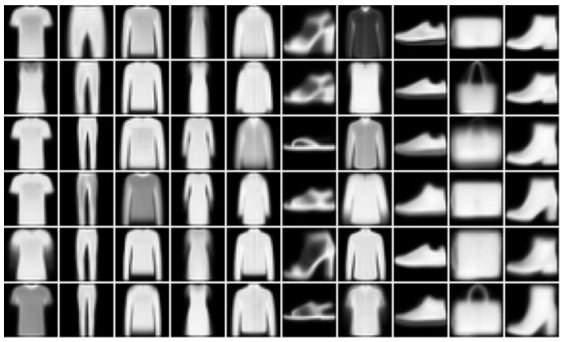


شکل (۱۸): تصاویر الگوهای مثبت حاصل از یادگیری توسط روش ED-WTA± . در هر ستون الگوهای مربوط به یک کلاس نمایش داده شده است

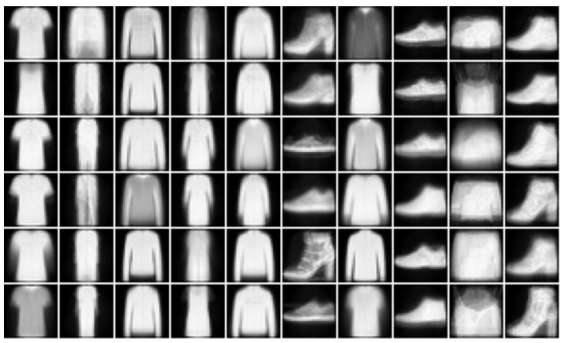


شکل (۱۹): تصاویر حاصل از یادگیری الگوهای منفی توسط روش ED-WTA± . در هر ستون الگوهای مربوط به یک کلاس نمایش داده شده است

در شکل (۲۰) و (۲۱) الگوهای بدسـت آمـده از روش میـانگین-گیری Mean-PNPL رسم شـده اسـت. ایـن شـکل نشـان مـی‌دهـد، الگوهای مثبت که میانگین نمونه‌های برنده شده توسط هر مرکز مـی-باشند، شباهت کمتری به الگوهای منفی حاصل از میـانگین داده‌هـای اشـــتباه دارنـــد. همچنـــین نتیجـــه‌ی روش mean-PNPL بـــر FashionMNIST دقت پایین‌تری نسبت به روش ED-WTA± دارد.



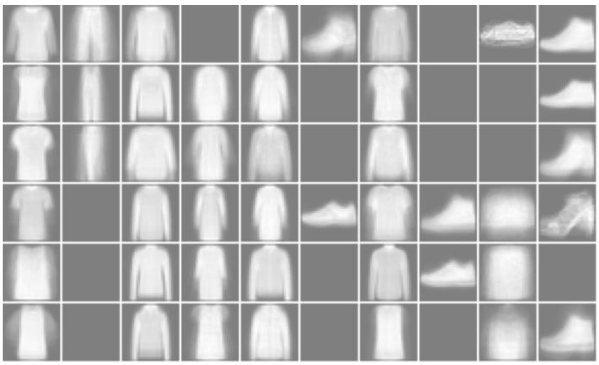
شکل(۱۹): الگوهای مثبت در روش Mean-PNPL



شکل (۲۰): الگوهای منفی حاصل در روش Mean-PNPL

اگر مقداردهی اولیه‌ی وزن نـورون‌هـای خروجـی در روش SVM چندالگویه با خوشه‌بندی Kmeans انجام شود، با آموزش این روش بر روی مجموعه دادگان Fashion، میانگین بردارهای پشتیبان بـه‌دسـت

آمده برای هر نورون خروجی شبکه بـه صـورت شـکل (۲۱) بـه‌دسـت می‌آید.



شکل(۲۱): مراکز حاصل از میانگین بردارهای پشتیبان به‌دست آمده از روش Multi-prototype SVM

مراجع

[1] Esmaeli H, Ghiasi-Shirazi K, Harati A. Online learning of positive and negative prototypes with explanations based on kernel expansion. Journal of Iranian Association of Electrical and Electronics Engineers 2023; 20 (1) :67-77

[2] R. Zarei-Sabzevar, K. Ghiasi-Shirazi, and A. Harati, "Prototype-based interpretation of the functionality of neurons in winner-take-all neural networks", IEEE Transactions on Neural Networks and Learning Systems, 2022.

[3] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning", Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017.

[4] O. Li, H. Liu, C. Chen, and C. Rudin, "Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions", In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. 2018.

[5] C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin, and J.K. Su, "This looks like that: deep learning for interpretable image recognition", Advances in neural information processing systems 32, 2019.

[6] G. Chen, T. Zhang, J. Lu, and J. Zhou, "Deep meta metric learning", In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9547-9556. 2019.

[7] V. Feldman, "Does learning require memorization? a short tale about a long tail", Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing. 2020.

[8] J. Bien, and R. Tibshirani, "Prototype selection for interpretable classification", The Annals of Applied Statistics: 2403-2424, 2011.

[9] S.Ö. Arik, and T. Pfister, "Protoattend: Attention-based prototypical learning", The Journal of Machine Learning Research, 21(1), pp.8691-8725, 2020.

[10] K. Chen, and C.G. Lee, "Incremental few-shot learning via vector quantization in deep embedded space", International Conference on Learning Representations. 2021.

[11] Q. Qian, L. Shang, B. Sun, J. Hu, H. Li, and R. Jin, "SoftTriple loss: Deep metric learning without triplet sampling", In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6450-6458. 2019.

[12] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering", Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[13] C. Cortes, and V. Vapnik, "Support-vector networks", Machine learning 20.3, pp. 273-297, 1995.

[14] G. Brown, M. Bun, V. Feldman, A. Smith, and K. Talwar, "When is memorization of irrelevant training data necessary for high-accuracy learning?", Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing. 2021.

[15] V. Feldman, and C. Zhang, "What neural networks memorize and why: Discovering the long tail via influence estimation", Advances in Neural Information Processing Systems 33, pp. 2881-2891, 2020.

[16] M.E. Mavroforakis, and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification", IEEE transactions on neural networks 17.3 pp. 671-682, 2006.

[17] C.K. Yeh, J. Kim, I.E.H. Yen, and P.K. Ravikumar, "Representer point selection for explaining deep neural networks", ,Advances in neural information processing systems 31, 2018.

[18] K. Allen, E. Shelhamer, H. Shin, and J. Tenenbaum, "Infinite mixture prototypes for few-shot learning", In International Conference on Machine Learning, pp. 232-241. PMLR, 2019.

[19] J. Chen, L.M. Zhan, X.M. Wu, and F.L. Chung, "Variational metric scaling for metric-based meta-learning", AAAI Conference on Artificial Intelligence, vol. 34, no. 04, pp. 3478–3485, 2020

[20] K. Crammer, and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines", Journal of machine learning research 2. Dec 265-292, 2001.

[21] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online passive aggressive algorithms", Journal of Machine Learning Research, 2006.

[22] F. Aiolli, A. Sperduti, and Y. Singer, "Multiclass Classification with Multi-Prototype Support Vector Machines", Journal of Machine Learning Research 6.5 2005.

[23] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms". Neural computation, 12(5), 1207-1245, 2000.

[24] M.E. Tipping, "Sparse Bayesian learning and the relevance vector machine", Journal of machine learning research 1. 211-244. 2001.

[25] http://ocw.um.ac.ir/streams/course/view/163.html

[26] S. Arora, H. Khandeparkar, M. Khodak, O. Plevrakis, and N. Saunshi, "A theoretical analysis of contrastive unsupervised representation learning", In International Conference on Machine Learning, pp. 5628-5637. PMLR, 2019.

[27] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D.Krishnan, "Supervised contrastive learning", Advances in Neural Information Processing Systems 33 ,18661-18673, 2020.

[28] K. Ghiasi-Shirazi, "Competitive cross-entropy loss: A study on training single-layer neural networks for solving nonlinearly separable classification problems", Neural Processing Letters, vol. 50, no. 2, pp. 1115–1122, 2019.

زیرنویس‌ها

[1] Support Vector Machines
[2] Convolutional Neural Networks
[3] Few-shot
[4] Learning vector quantization
[5] ±Euclidean Distance-based Winner-Take-All Networks
[6] Softmax
[7] Hard Samples
[8] Hardness-aware prototype-mining
[9] Support Vector Machine
[10] Attention Mechanism
[11] Relational Attention
[12] explainable
[13] Embedding space
[14] Distance-based metric-learning
[15] Representer theorem
[16] Competitive Cross-Entropy
[17] Mean-Positive Negative Prototype Learning
[18] Few-shot Learning