

# مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی با نظارت در چارچوب کاهش بعد بدون اتلاف برای شناسایی چهره

سمیه احمدخانی و پیمان ادبی

گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران

## چکیده

در این مقاله ابتدا مدل بانظارت روش ترکیبی تحلیل مؤلفه اصلی احتمالاتی<sup>۱</sup> (SPPCAMM) ارائه شده است؛ سپس با درنظرگرفتن جریمه نگاشت در یادگیری مدل پیش‌گو، روشی برای شناسایی چهره با استفاده از یک رویکرد کاهش بعد بدون اتلاف ارائه شده است. در روش پیشنهادی ابتدا یک خمینه زیربنایی محلی خطی با استفاده از مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظارت از نمونه داده‌ها به دست می‌آید؛ سپس دسته‌بند ماشین بردار پشتیبان با اعمال جریمه نگاشت به عنوان مدل پیش‌گوی مذکور با استفاده از این خمینه محلی خطی آموزش داده می‌شود. بدین ترتیب از مزایای کاهش بعد در مدل پیش‌گو استفاده می‌شود؛ و در عین حال جلوی از دست رفتن اطلاعات مفید گرفته می‌شود. برای آموزش و ارزیابی روش پیشنهادی، از پایگاه داده‌های چهره شناخته شده است. روش استخراج ویژگی گابور بر روی تصاویر به کار گرفته شده است. نتایج آزمایش‌ها نشان می‌دهد که روش پیشنهادی نسبت به بسیاری از روش‌های معمول که کاهش بعد را انجام داده و سپس دسته‌بند را آموزش می‌دهند، و همچنین نسبت به روش جریمه نگاشت مبتنی بر مدل‌های کاهش بعد خطی و غیرخطی، دقت بیشتری دارد.

وازگان کلیدی: کاهش بعد بدون اتلاف، مدل ترکیبی، تحلیل مؤلفه اصلی احتمالی، بانظارت، جریمه نگاشت.

یکی از روش‌های کاهش بعد محبوب روش تحلیل مؤلفه اصلی<sup>۲</sup> (PCA) است، که تجزیه مقادیر منفرد<sup>۳</sup> (SVD) را روی ماتریس داده اجرا کرده و زیرفضایی با استفاده از مقادیر ویژه بزرگ ایجاد می‌کند (ون در ماتن و همکاران، ۲۰۰۹). بسیاری از روش‌های کاهش بعد بدون نظارت هستند؛ به طوری که تنها روی داده‌های بدون برچسب مشاهده شده ورودی تمرکز می‌کنند؛ اما در موافقی که مقادیر برچسب‌ها یا خروجی در دسترس است، وارد کردن این اطلاعات در نگاشت و رسیدن به یک افکنش بانظارت برای داده ورودی در کاربردهای مانند دسته‌بندی می‌تواند مفیدتر باشد. بنابراین روش‌های کاهش بعد بانظارت بهمنظور به کارگیری اطلاعات خروجی ارائه شده‌اند (یوسو و همکاران، ۲۰۰۶).

## ۱- مقدمه

یادگیری یک مدل دسته‌بندی در فضایی با بعد بالا و با وجود محدودیت در تعداد نمونه‌های آموزشی، کار دشواری است. از این‌رو روش‌های کاهش بعد فضای ویژگی به‌منظور بهسازی یادگیری در این مدل‌ها مورد توجه پژوهش‌گران قرار گرفته‌اند. روش‌های کاهش بعد علاوه‌بر کوچک‌کردن بردار ویژگی، به‌نوعی باعث افزایش دقت دسته‌بندی نیز می‌شوند. از آنجا که بعد فضای ورودی در بسیاری از مسائل بیش از حد لازم، بزرگ است، استفاده از روش‌های کاهش بعد علاوه‌بر کم‌کردن طول بردار ویژگی و کاهش هزینه محاسباتی مراحل بعدی، بالابردن کارایی از طریق مقابله با معضل بعد<sup>۱</sup> را نیز در پی خواهد داشت. به همین دلیل این روش‌ها در عمل دارای اهمیت بوده و به طور وسیع مورد مطالعه قرار گرفته‌اند (بیشاپ، ۲۰۰۶).

<sup>2</sup> Principal Component Analysis

<sup>3</sup> Singular Value Decomposition (SVD)

<sup>۱</sup> Curse of Dimensionality

هسته<sup>۷</sup> ارائه شده است (ژو و مارتینز، ۲۰۰۶). در این روش با استفاده از تقسیم هر دسته به تعدادی زیردسته سعی می‌شود تا بر فرض گاسی تک‌قله‌ای بودن تمام دسته‌ها غلبه شود؛ اما این روش مجبور است هر دسته را به کمک روش‌های خوشه‌بندی به زیرگروه‌های تقسیم کند. تحلیل تمايز زیردسته ترکیبی با گام کسری<sup>۸</sup> روش بانظار دیگری برای کاهش بعد محسوب می‌شود (گاللیس و همکاران، ۲۰۱۳)، که در آن مانند تحلیل تمايز زیردسته هر دسته به زیردسته‌هایی تقسیم می‌شود. ویژگی مهم این روش حل مشکل جدایی زیردسته‌ها است که زمانی رخ می‌دهد که رتبه ماتریس فاصله داخلی درون زیردسته‌ای از تعداد بعد کاهش‌یافته بیشتر باشد. نسخه مبتنی بر هسته این روش در (گاللیس و همکاران، ۲۰۱۳) ارائه شده است. در (هان و همکاران، ۲۰۱۳) نیز روشی برای حل مسئله کاهش بعد بانظار ارائه شده است.

در میان روش‌های کاهش بعد ارائه شده تاکنون روش‌های بدون نظارت بسیاری وجود دارد؛ و بهدلیل این که بانظار کردن یک روش می‌تواند نتایج را در کاربردهای مانند دسته‌بندی بهبود دهد، پژوهش‌هایی در زمینه بانظار کردن برخی روش‌های کاهش بعد بدون نظارت صورت گرفته است. در (لی و گو، ۲۰۰۶) نسخه‌ای بانظار روش Isomap به نام SE-Isomap پیشنهاد شده است. در این روش ماتریس فاصله ژئودسیک<sup>۹</sup> هر دسته با توجه به اطلاعات مربوط به آن دسته به دست می‌آید؛ و سپس روش مقیاس‌سازی چندبعدی<sup>۱۰</sup> (MDS) بر روی گرافی که با توجه به این ماتریس‌های فاصله به دست آمده، اعمال می‌شود. در (چن و لیو، ۲۰۱۱) نیز پیشنهادهایی برای بانظار کردن روش جانشانی محلی خطی<sup>۱۱</sup> (LLE) مطرح شده است.

در این مقاله ابتدا مدل بانظار روش ترکیبی تحلیل مؤلفه اصلی احتمالاتی<sup>۱۲</sup> (PPCAMM) ارائه شده که SPPCAMM نامیده می‌شود. این مدل توسعه‌ای از مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی است که اطلاعات مربوط به برچسب داده را در نگاشت یادگیری شده منظور می‌کند؛ سپس این رویکرد بانظار کاهش بعد محلی خطی SPPCAMM در چارچوبی به نام جریمه نگاشت<sup>۱۳</sup> (زنگ و

روش تحلیل تمايز خطی<sup>۱</sup> یا LDA روشی بانظار است که به کمک یک بهینه‌سازی ساده بردار افکنش را طوری به دست می‌آورد که فاصله درون دسته‌ای نمونه‌های آموزشی کمینه و فاصله بین دسته‌ای آن‌ها بیشینه شود (فیشر، ۱۹۳۸). روش تحلیل تمايز خطی وقتی به خوبی کار می‌کند که نمونه داده‌ها جدایی‌پذیر خطی باشند. برای غلبه بر این محدودیت روش تحلیل تمايز هسته<sup>۲</sup> (KDA) ارائه شده است (میکا و همکاران، ۱۹۹۹). فرض دیگری در روش تحلیل تمايز نخستین وجود دارد که توزیع احتمال هر دسته را محدود به توزیع گاسی تک‌قله‌ای می‌کند (فیشر، ۱۹۳۸). روش تحلیل فیشر حاشیه‌ای<sup>۳</sup> (MFA) با استفاده از ساختار جانشانی گراف<sup>۴</sup> (یان و همکاران، ۲۰۰۷)، سعی در غلبه بر محدودیت‌های روش LDA دارد. در مقایسه با LDA، روش MFA مزایای زیر را دارد: ۱) جهت‌های افکنش در دسترس خیلی بیشتر از LDA است، و ۲) هج فرضی روی توزیع داده هر دسته وجود ندارد (یان و همکاران، ۲۰۰۷). به تازگی روش حفظ همسایگی و جانشانی تمايز حاشیه‌ای<sup>۵</sup> (NP-MDE) بر اساس جانشانی خطی گراف و روش MFA پیشنهاد شده است؛ که می‌تواند پراکندگی درون دسته‌ای را کمینه و هم‌زمان حاشیه بین دسته‌ای را بیشینه کردد، و به علاوه ساختار همسایگی با هر دسته را نیز حفظ کند (لان و همکاران، ۲۰۱۲).<sup>۶</sup>

روش TRIMAP<sup>۷</sup> روش بانظار دیگری برای کاهش بعد است (چن و همکاران، ۲۰۱۰). این روش که ترکیبی از روش Isomap (ژو و مارتینز، ۲۰۰۶) و تحلیل تمايز خطی است، ابتدا از Isomap برای پیدا کردن زیرفضای مناسب استفاده می‌کند؛ اما این روش تابع فاصله اقلیدسی معمولی را برای تعیین فواصل به کار نمی‌برد؛ بلکه از تابع فاصله‌ای استفاده می‌کند که فاصله نمونه‌های متعلق به دسته‌های متفاوت را بیشتر از حد واقعی تخمین می‌زند. این کار باعث می‌شود تمايز بیشتری بین دسته‌های متفاوت ایجاد شود؛ سپس این مدل از تحلیل تمايز خطی برای ایجاد تمايز بیشتر میان دسته‌ها استفاده می‌کند.

برای رفع هم‌زمان مشکلات منتج از فرض‌های روش تحلیل تمايز خطی، روش تحلیل تمايز زیردسته مبتنی بر

<sup>7</sup> Kernel Sub-class Discriminant Analysis (KSDA)

<sup>8</sup> Fractional step Mixture Subclass Discriminant Analysis (FMSDA)

<sup>9</sup> Geodesic

<sup>10</sup> Multidimensional Scaling

<sup>11</sup> Locally Linear Embedding

<sup>12</sup> Probabilistic Principal Component Analysis

<sup>13</sup> Projection penalties

<sup>1</sup> Linear Discriminant Analysis (LDA)

<sup>2</sup> Kernel Discriminant Analysis (KDA)

<sup>3</sup> Marginal Fisher Analysis (MFA)

<sup>4</sup> Graph Embedding

<sup>5</sup> Neighborhood Preserving and Marginal Discriminant Embedding

<sup>6</sup> Tensor-based Riemannian Manifold Distance-Approximating Projection (TRIMAP)



ابعاد  $N \times N$  و  $M \times M$  بوده و **D** ماتریسی قطری متشكل از مقادیر منفرد می‌باشد، که به ترتیب نزولی مرتب شده و در امتداد قطر قرار گرفته‌اند. ماتریس متشكل از  $K$  ستون نخست از ماتریس **U** که با  $K$  نشان داده می‌شود، به عنوان نگاشت  $\Psi$  در نظر گرفته می‌شود (پیرسون، ۱۹۰۱).

روش تحلیل مؤلفه اصلی احتمالاتی (PPCA)، یک بیان احتمالاتی از روش PCA را ارائه می‌کند (تپینگ و بیشап، ۱۹۹۹-ب). روش PPCA یک مدل متغیر نهان<sup>۵</sup> است و فرآیند تولید هر ورودی  $x$  را به صورت  $x = W_x z + \mu_x + \epsilon_x$  مدل می‌کند، که  $z \in R^K$  متغیرهای نهان و  $W_x$  ماتریسی  $M \times K$  است، که factor loading نامیده می‌شود. در این مدل احتمالاتی به طور معمول فرض می‌شود که بردار متغیرهای نهان  $z$  دارای توزیع گاسین با میانگین صفر و ماتریس کواریانس واحد است (یعنی متقاضان شعاعی به صورت  $(z \sim N(\mathbf{0}, \mathbf{I})$ )، و  $\epsilon_x$  نوافه می‌باشد که دارای توزیع گاسین پارامترهای مدل با یک روش تکراری به نام الگوریتم میانگین گیری-بیشینه‌سازی<sup>۶</sup> EM تعیین می‌شوند. لگاریتم شباهت<sup>۷</sup> برای نمونه داده‌ها در این مدل به صورت زیر تعریف می‌شود (تپینگ و بیشап، ۱۹۹۹-الف):

$$l = \sum_{n=1}^N \ln\{p(x_n)\} \quad (1)$$

که در آن  $p(x_n)$  توزیع حاشیه‌ای  $X$  است، و به صورت  $p(x_n) = \int p(x|z)p(z)dz$  تعریف می‌شود. نتایج مرتبط می‌تواند در قالب قضیه ۱ به صورت زیر خلاصه شود. جزیيات اثبات این قضیه در مرجع (یو و همکاران، ۲۰۰۶) و (تپینگ و بیشап، ۱۹۹۹-الف) قابل دستیابی است:

قضیه ۱: اگر  $S_x = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_n)^T (x_n - \mu_n)$  ماتریس کواریانس نمونه داده‌های  $\{x_n\}_{n=1}^N$  باشد و  $\lambda_1 \geq \dots \geq \lambda_M$  مقادیر ویژه متناظر با بردارهای ویژه  $u_1, \dots, u_M$  ماتریس  $S_x$  باشد، و فضای نهان در مدل PPCA، فضایی  $K$  بعدی باشد، آن‌گاه (یو و همکاران، ۲۰۰۶):

(الف) تخمین بیشترین شباهت برای  $W_x$  به صورت زیر به دست می‌آید:

$$W_x = U_K (\Lambda_K - \delta_x^2 I)^{\frac{1}{2}} R \quad (2)$$

اشنايدر، ۲۰۱۰) به کار گرفته می‌شود. جریمه نگاشت این امکان را فراهم می‌آورد که از کاهش بعد برای بهبود دقیق مدل پیش‌گو بدون خطر از دستدادن اطلاعات مرتبط استفاده کنیم (زانگ و اشنايدر، ۲۰۱۰). چنان که در بخش ۴ خواهیم دید، نتایج آزمایش‌ها بر روی پایگاه داده‌های تصاویر چهره، نشان‌دهنده کارایی بالاتر روش پیشنهادی SPPCAMM در چهارچوب جریمه نگاشت نسبت به سایر روش‌های بررسی شده است.

ساختار مقاله به صورت زیر است: در بخش ۲ مژوری بر کارهای مرتبط خواهیم داشت، و چند روش کاهش بعد بررسی می‌شود. در بخش ۳ روش پیشنهادی یعنی مدل با نظارت روش ترکیبی تحلیل مؤلفه اصلی احتمالاتی و به کارگیری آن در چارچوب جریمه نگاشت معروفی می‌شود. نتایج پیاده‌سازی‌ها در بخش ۴ گزارش می‌شوند. درنهایت تحلیل نتایج در بخش ۵، و جمع‌بندی مقاله در بخش ۶ ارائه خواهد شد.

## ۲- مژوری بر کارهای مرتبط

در این بخش روش‌های کاهش بعد مرتبط با روش پیشنهادی را مورد بررسی قرار می‌دهیم. این روش‌ها عبارت از روش تحلیل مؤلفه اصلی<sup>۱</sup> (PCA)، تحلیل مؤلفه اصلی احتمالاتی<sup>۲</sup> (PPCA)، تحلیل مؤلفه اصلی احتمالاتی با نظارت<sup>۳</sup> (SPPCA)، و مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی<sup>۴</sup> (PPCAMM) (تپینگ و بیشап، ۱۹۹۹-الف) هستند. در ادامه فرض می‌کنیم مجموعه‌ای از  $N$  نمونه آموزشی در اختیار است. ورودی  $u_m$  که با  $x_n \in X$  نمایش داده می‌شود، توسط یک بردار ویژگی  $M$  بعدی مشخص می‌شود ( $x_n \in X \subseteq R^M$ ). برای کاهش بعد، هدف یافتن نگاشتی به صورت  $\Psi: X \rightarrow Z$  است که هر ورودی را به یک فضای  $K$  بعدی با شرط  $Z \subseteq R^K$  نگاشت می‌کند (Z).

روش PCA یک روش کاهش بعد بدون نظارت است که هدف آن یافتن مؤلفه‌های اصلی، که جهت‌هایی با بیشترین واریانس را برای داده مشخص می‌کنند، است. اگر فرض کنیم  $X = [x_1, \dots, x_N]^T$  ماتریس ورودی متمرکز شده را مشخص کند، که در آن میانگین نمونه‌ها از هر نمونه کم شده است،  $X = VDU^T$  تجزیه مقدار منفرد  $X$  را مشخص می‌کند که  $V$  و  $U$  ماتریس‌های معتمد ستونی با

<sup>1</sup> Principal Component Analysis

<sup>2</sup> Probabilistic Principal Component Analysis

<sup>3</sup> Supervised Probabilistic Principal Component Analysis

<sup>4</sup> PPCA Mixture Model

کرد (جزییات این قضیه در مرجع (یو و همکاران، ۲۰۰۶) موجود است):

قضیه ۲: ماتریس کواریانس نمونه داده‌های هنجارسازی شده  $\{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$  را به صورت زیر تعریف می‌کنیم:

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N \Phi^{-\frac{1}{2}} \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix} \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix}^T \quad \Phi^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\delta_x^2} \mathbf{S}_x & \frac{1}{\delta_x \delta_y} \mathbf{S}_{xy} \\ \frac{1}{\delta_x \delta_y} \mathbf{S}_{yx} & \frac{1}{\delta_y^2} \mathbf{S}_y \end{pmatrix} \quad (6)$$

و  $\lambda_1 \geq \dots \geq \lambda_{(M+L)}$  مقادیر ویژه متناظر با بردارهای ویژه  $\mathbf{u}_{1, \dots, M+L}$  از ماتریس  $\mathbf{S}$  می‌باشند. ماتریس  $\mathbf{S}_x$  ماتریس کواریانس بین ورودی‌ها،  $\mathbf{S}_y$  ماتریس کواریانس بین خروجی‌ها، و  $\mathbf{S}_{xy}$  ماتریس کواریانس بین ورودی‌ها و خروجی‌ها است. اگر فضای نهان را  $K$  بعدی در نظر بگیریم، نتایج زیر را خواهیم داشت (یو و همکاران، ۲۰۰۶):

(الف) در SPPCA، ماتریس‌های  $\mathbf{W}_x$  و  $\mathbf{W}_y$  به صورت زیر به دست می‌آیند:

$$\mathbf{W}_x = \frac{1}{\delta_x} \mathbf{U}_x (\Lambda_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R} \quad (7)$$

$$\mathbf{W}_y = \frac{1}{\delta_y} \mathbf{U}_y (\Lambda_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R}$$

که  $\mathbf{U}_x$  شامل  $M$  سطر نخست از ماتریس  $[\mathbf{u}_1, \dots, \mathbf{u}_K]$  بوده و  $\mathbf{U}_y$  سطر باقیمانده از آن است. ماتریس  $\mathbf{R}$  یک ماتریس  $K \times K$  متعامد دلخواه است.

(ب) نتیجه نگاشت برای ورودی متغیر کشیده  $\mathbf{x}^*$ ، بردار  $\mathbf{z}^*$  است که به صورت زیر به دست می‌آید:

$$\mathbf{z}^* = \frac{1}{\delta_x} \mathbf{R}^T (\Lambda_K - \mathbf{I})^{-\frac{1}{2}} [\mathbf{U}_x^T \mathbf{U}_x + (\Lambda_K - \mathbf{I})^{-1}]^{-1} \mathbf{U}_x^T \mathbf{x}^* \quad (8)$$

در حالت خاص اگر  $L=0$  باشد، آن‌گاه مدل بدون نظارت است و  $\mathbf{S} = \frac{1}{\delta_x^2} \mathbf{S}_x$  خواهد بود، که مشابه ماتریس مربوط به PPCA خواهد شد (یو و همکاران، ۲۰۰۶).

در مدل ترکیبی تحلیل مؤلفه‌های اصلی احتمالاتی (PPCAMM) از ترکیب خطی این مدل‌ها بررسی مجموعه داده‌ها استفاده می‌شود. درواقع روش PPCAMM چند PPCA محلی است که در این روش بخش‌بندی مناسب نمونه داده‌ها و تعیین مؤلفه‌های اصلی مدل برای هر بخش به طور همزمان با استفاده از تخمین بیشترین شباهت تعیین می‌شود. لگاریتم شباهت نمونه داده‌ها برای این مدل ترکیبی به صورت زیر تعیین می‌شود:

که  $(\mathbf{u}_1, \dots, \mathbf{u}_K) = \text{diag}(\lambda_1, \dots, \lambda_K)$  و  $\mathbf{R}$  یک ماتریس متعامد  $K \times K$  است.

(ب) نتیجه نگاشت برای ورودی جدید  $\mathbf{x}^*$ ، بردار  $\mathbf{z}^*$  است که به صورت زیر به دست می‌آید:

$$\mathbf{z}^* = \mathbf{R}^T (\Lambda_K - \delta_x^2 \mathbf{I})^{\frac{1}{2}} \Lambda_K^{-1} \mathbf{U}_K^T (\mathbf{x}^* - \boldsymbol{\mu}) \quad (3)$$

مدل PPCA به دلیل بیان در چارچوب احتمالاتی، مزایایی نسبت به PCA دارد، که از آن جمله می‌توان رویه یادگیری سریع الگوریتم EM، ارائه روشی اصولی برای مدیریت ورودی‌های مفقوصه، و امکان در نظر گرفتن مدل ترکیبی برای PCA را بشمرد (تپینگ و بیشاب، ۱۹۹۹-۱۹۹۹).

(الف) (تپینگ و بیشاب، ۱۹۹۹-۱۹۹۹-۱-۲).

در مسائل یادگیری بانتظارت هر داده  $\mathbf{x}$  با یک مقدار خروجی  $\mathbf{y} \in \mathbb{Y}$  متناظر است. مدل SPPCA نسخه بانتظارت مدل PPCA است، که در آن برای مسایل رگرسیون  $\mathbf{y} \in \mathbb{R}$  و برای مسایل دسته‌بندی  $\{\mathbf{y} \in \{-1, +1\}\}$  در نظر گرفته می‌شود (یو و همکاران، ۲۰۰۶). مدل SPPCA بر این باور است که بین فضای ورودی  $X$  و فضای خروجی  $Y$  کواریانس وجود دارد و این امکان وجود دارد که PPCA را به مدلی که این کواریانس را به خوبی نشان می‌دهد تعمیم داد. بیان ریاضی روشن SPPCA به صورت زیر است: اگر بعد فضای خروجی را  $L$  فرض کنیم، هر ورودی  $\mathbf{x}$  متناظر با یک بردار خروجی مشاهده شده  $\mathbf{y} = [y_1, \dots, y_L]^T \in \mathbb{R}^L$  است. در  $\mathbf{y} = \mathbf{W}_x \mathbf{x} + \boldsymbol{\mu}_x + \boldsymbol{\varepsilon}_x$  توسط مدل متغیر نهان زیر تولید می‌شود:

$$\mathbf{y} = \mathbf{W}_y \mathbf{z} + \boldsymbol{\mu}_y + \boldsymbol{\varepsilon}_y \quad (4)$$

متغیرهای نهان  $\mathbf{z} \sim N(0, I)$  بین  $\mathbf{x}$  و  $\mathbf{y}$  مشترک

می‌باشند، و دو مدل نوفر که از یکدیگر مستقل هستند و هر دو دارای توزیع ایزوتوپیک گاسین می‌باشند، به صورت  $\boldsymbol{\varepsilon}_x \sim N(\mathbf{0}, \delta_x^2 \mathbf{I})$  و  $\boldsymbol{\varepsilon}_y \sim N(\mathbf{0}, \delta_y^2 \mathbf{I})$  در نظر گرفته می‌شوند.

در این روش اگر قرار دهیم:

$$\boldsymbol{\Phi} = \begin{pmatrix} \delta_x^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \delta_y^2 \mathbf{I} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{pmatrix} \quad (5)$$

بر اساس فرضیاتی که برای این روش در نظر گرفته شده است، به راحتی دیده می‌شود که  $(\mathbf{x}, \mathbf{y})$  دارای توزیع توأم گاسین با میانگین  $\boldsymbol{\mu}$  و ماتریس کواریانس  $\boldsymbol{\Phi} + \mathbf{W} \mathbf{W}^T$  هستند (یو و همکاران، ۲۰۰۶). نتایج مرتبط با این روش را می‌توان به طور خلاصه در قالب قضیه ۲ به شکل زیر بیان

فصل نهم



همان طور که ذکر شد، تپینگ و بیشاب روشی مطرح کردند که در آن PCA در چارچوب بیشترین شباهت<sup>۱</sup> و براساس صورت خاصی از مدل متغیر نهان گاسی فرمول بندی شده است. این روش که به عنوان روش تحلیل مؤلفه اصلی احتمالاتی (PPCA) معروفی شده است، منجربه ایجاد یک مدل ترکیبی خوش تعریف برای تحلیل مؤلفه اصلی احتمالاتی محلی خطی می شود، که پارامترهای آن می توانند با استفاده از الگوریتم تکراری EM تعیین شوند (تپینگ و بیشاب، ۱۹۹۹-الف)، اما این روش هنگامی که خروجی های مربوط به نمونه داده های ورودی در اختیار باشد، مانند مسائل دسته بندی و رگرسیون، قادر به استفاده مفید از این اطلاعات نیست.

با توجه به اهمیت مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی در فضاهای غیر خطی و پیچیده و نیز غلبه بر محدودیت های روش PCA و نیز با توجه به این موضوع که روش های کاهش بعد بانتظارت در مسائل مختلف به خصوص دسته بندی و پیش بینی، به دلیل استفاده از اطلاعات مسئله از کارایی بهتری برخوردار هستند، ارائه روش بانظارت مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی در این مقاله مورد توجه قرار گرفته است. روش ارائه شده تحت عنوان SPPCAMM معرفی می شود. در روش پیشنهادی از ترکیب خطی مدل ها بر روی مجموعه داده ها با در نظر گرفتن برچسب دسته ها استفاده می شود.

روش SPPCAMM را می توان ترکیب چند PCA محلی دانست، که در آن بخش بندی مناسب نمونه داده ها و تعیین مؤلفه های اصلی مدل برای هر بخش به طور همزمان با استفاده از تخمین بیشترین شباهت انجام می شود. لگاریتم شباهت نمونه داده ها برای این مدل مشابه (تپینگ و بیشاب، ۱۹۹۹-الف) بوده و به صورت زیر تعریف می شود، با این تفاوت که در این حالت هر نمونه به صورت  $(\mathbf{x}_n, \mathbf{y}_n)$  است:

$$l = \sum_{n=1}^N \ln\{p(\mathbf{x}_n, \mathbf{y}_n)\} = \sum_{n=1}^N \ln\{\sum_{i=1}^M \pi_i p(\mathbf{x}_n, \mathbf{y}_n | i)\} \quad (14)$$

که  $M$  تعداد مدل های ترکیبی،  $N$  تعداد نمونه داده ها و  $p(\mathbf{x}_n, \mathbf{y}_n | i)$  یک مدل SPPCA است. اگر پاسخ احتمال پسین مدل نام برای تولید نقطه داده  $(\mathbf{x}_n, \mathbf{y}_n)$  را به صورت  $\mathbf{R}_{ni} = p(i | \mathbf{x}_n, \mathbf{y}_n)$  بیان کنیم، از رابطه زیر می توان این احتمال را محاسبه کرد:

$$\mathbf{R}_{ni} = \frac{p(\mathbf{x}_n, \mathbf{y}_n | i) \pi_i}{p(\mathbf{x}_n, \mathbf{y}_n)} \quad (15)$$

<sup>2</sup> Maximum likelihood

$l = \sum_{n=1}^N \ln\{p(\mathbf{x}_n)\} = \sum_{n=1}^N \ln\{\sum_{i=1}^M \pi_i p(\mathbf{x}_n | i)\}$  (۹)  
که  $M$  تعداد مدل های مشارکت کننده در ترکیب،  $N$  تعداد نمونه داده ها، و  $p(\mathbf{x}_n | i)$  یک مدل PPCA است. ضریب  $\pi_i$  نسبت مشارکت مؤلفه مدل نام در ترکیب مورد نظر می باشد، که برای آن محدودیت های احتمالاتی  $0 \leq \pi_i \leq 1$  برقرار است. هر مؤلفه از ترکیب بردار میانگین  $\mu_i$  و همچنین پارامترهای  $\mathbf{W}_{xi}$  و  $\delta_{xi}$  مربوط به خود را دارد (تپینگ و بیشاب، ۱۹۹۹-الف).

می توان برای بهینه کردن پارامترهای هر مؤلفه از ترکیب الگوریتم تکراری EM را مورد استفاده قرار داد. نشان داده می شود  $\mathbf{R}_{ni} = p(i | \mathbf{x}_n)$ ، که پاسخ احتمال پسین مؤلفه نام برای تولید نقطه داده  $\mathbf{x}_n$  را بیان می کند، به صورت زیر به دست می آید (تپینگ و بیشاب، ۱۹۹۹-الف):

$$\mathbf{R}_{ni} = \frac{p(\mathbf{x}_n | i) \pi_i}{p(\mathbf{x}_n)} \quad (10)$$

همچنین می توان نشان داد که  $\pi_i$  و  $\mu_{xi}$  به صورت زیر به دست می آیند (تپینگ و بیشاب، ۱۹۹۹-الف):

$$\tilde{\pi}_i = \frac{1}{N} \sum_{n=1}^N \mathbf{R}_{ni} \quad (11)$$

$$\tilde{\mu}_{xi} = \frac{\sum_{n=1}^N \mathbf{R}_{ni} \mathbf{x}_n}{\sum_{n=1}^N \mathbf{R}_{ni}} \quad (12)$$

به علاوه،  $\mathbf{W}_{xi}$  و  $\delta_{xi}^2$  با استفاده از بردارهای ویژه و مقادیر ویژه ماتریس کواریانس وزن دار محلی مدل نام که به صورت زیر تعریف می شود، به دست می آیند (تپینگ و بیشاب، ۱۹۹۹-الف):

$$\mathbf{S}_{xi} = \frac{1}{\tilde{\pi}_i N} \sum_{n=1}^N \mathbf{R}_{ni} (\mathbf{x}_n - \tilde{\mu}_{xi}) (\mathbf{x}_n - \tilde{\mu}_{xi})^T \quad (13)$$

### ۳- روش پیشنهادی

#### ۱-۱- مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظارت

روش PCA با وجود محدودیتی که به خاطر خطی بودن دارد، یکی از پر کاربرد ترین روش های کاهش بعد است. تعمیم های مختلفی از روش PCA برای غلبه بر محدودیت خطی بودن آن مطرح شده است. یک پیشنهاد مطرح این است که مجموعه داده های پیچیده به صورت ترکیبی خطی محلی از نگاشته های PCA بیان شود. با این وجود PCA متناظر با یک چگالی احتمالاتی نیست و لذا روش منحصر به فردی برای ترکیب مدل های PCA وجود ندارد. پیشنهادهای قبلی برای توسعه یک مدل ترکیبی<sup>۱</sup> برای PCA تک منظوره بوده و عمومیت نداشته اند.

<sup>1</sup> Mixture model

رابطه بالا بردار ویژگی کاهش بعدیافته برای ورودی  $\mathbf{x}^*$  به کمک مؤلفه آن از ترکیب را نشان می‌دهد. بردار ویژگی کاهش بعدیافته نهایی به صورت زیر به دست می‌آید.  
ابتدا بردار ویژگی کاهش بعدیافته با هر مؤلفه از ترکیب را با استفاده از رابطه (۲۱) به دست می‌آوریم؛ سپس فاصله اقلیدسی بازیابی بردار کاهش بعدیافته از بردار اصلی را به دست می‌آوریم. مؤلفه‌ای که کمترین فاصله اقلیدسی را به دست دهد پیدا شده و بردار ویژگی کاهش بعدیافته نهایی با افکنش بردار اصلی بر آن مؤلفه به دست می‌آید.  
محاسبه مقادیر  $\delta_{xi}^2$  و  $\delta_{yi}^2$  با استفاده از روابط زیر انجام می‌شود (یو و همکاران، ۲۰۰۶) (تپینگ و بیشپ، ۱۹۹۹-الف):

$$\begin{aligned}\delta_{xi}^2 &= \frac{1}{M \sum_n \mathbf{R}_{ni}} \left\{ \sum_{n=1}^N \mathbf{R}_{ni} \left\| \mathbf{x}_n - \boldsymbol{\mu}_{xi} \right\|^2 + \mathbf{R}_{ni} \text{tr}(\mathbf{W}_x^T \mathbf{W}_x \mathbf{C}) - 2 \mathbf{R}_{ni} \text{tr}(\mathbf{X} \mathbf{W}_x \mathbf{Z}^T) \right\} \\ \delta_{yi}^2 &= \frac{1}{M \sum_n \mathbf{R}_{ni}} \left\{ \sum_{n=1}^N \mathbf{R}_{ni} \left\| \mathbf{y}_n - \boldsymbol{\mu}_{yi} \right\|^2 + \mathbf{R}_{ni} \text{tr}(\mathbf{W}_y^T \mathbf{W}_y \mathbf{C}) - 2 \mathbf{R}_{ni} \text{tr}(\mathbf{X} \mathbf{W}_y \mathbf{Z}^T) \right\}\end{aligned}\quad (22)$$

در روابط بالا،  $\mathbf{Z} = [\langle \mathbf{z}_1 \rangle, \dots, \langle \mathbf{z}_N \rangle]^T$ ،  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$  است، که  $\langle \cdot \rangle$  بیان کننده امید ریاضی بوده، و آماره‌های  $\langle \mathbf{z}_n \rangle$  و  $\langle \mathbf{z}_n \mathbf{z}_n^T \rangle$  با تخمین توزیع برای مشاهدات  $(\mathbf{x}_n, \mathbf{y}_n)$  به دست می‌آیند. وقتی که  $\mathbf{z}$  ورودی  $\mathbf{x}$  و خروجی  $\mathbf{y}$  هر دو در دست باشند، توزیع پسین برای  $(\mathbf{x}, \mathbf{y})$  داده شده به صورت زیر تعریف می‌شود (یو و همکاران، ۲۰۰۶):

$p(\mathbf{z}|\mathbf{x}, \mathbf{y}) \propto p(\mathbf{x}, \mathbf{y}|\mathbf{z})p(\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})p(\mathbf{z})$  (۲۳)  
با توجه به این که سه جمله سمت راست دارای توزیع گاسین هستند، توزیع پسین  $\mathbf{z}$  نیز گاسین ( $N(\boldsymbol{\mu}_z, \Sigma_z)$  خواهد بود، و داریم:

$$\begin{aligned}\boldsymbol{\mu}_z &= \mathbf{A}^{-1} \left[ \frac{1}{\delta_x^2} \mathbf{W}_x^T (\mathbf{x} - \boldsymbol{\mu}_x) + \frac{1}{\delta_y^2} \mathbf{W}_y^T (\mathbf{y} - \boldsymbol{\mu}_y) \right] \\ \Sigma_z &= \mathbf{A}^{-1} \\ \mathbf{A} &= \frac{1}{\delta_x^2} \mathbf{W}_x^T \mathbf{W}_x + \frac{1}{\delta_y^2} \mathbf{W}_y^T \mathbf{W}_y + \mathbf{I}\end{aligned}\quad (24)$$

بنابراین برای هر مدل در SPPCAMM نیز مشابه بالا داریم:  
 $\mathbf{A}_i = \frac{1}{\delta_{xi}^2} \mathbf{W}_{xi}^T \mathbf{W}_{xi} + \frac{1}{\delta_{yi}^2} \mathbf{W}_{yi}^T \mathbf{W}_{yi} + \mathbf{I}$  (۲۵)  
و آماره‌های  $\langle \mathbf{z}_n \mathbf{z}_n^T \rangle$  و  $\langle \mathbf{z}_n \rangle$  با استفاده از روابط بالا به صورت زیر تعریف می‌شوند:

که  $\pi_i$  نسبت ترکیب مدل نام مورد نظر با محدودیت‌های  $\sum_i \pi_i = 1$  است، و به صورت زیر به دست می‌آید:

$$\tilde{\pi}_i = \frac{1}{N} \sum_{n=1}^N \mathbf{R}_{ni} \quad (16)$$

رابطه بالا در گام M از الگوریتم EM و با بیشینه‌کردن لگاریتم شباهت داده نسبت به  $\pi_i$  به دست می‌آید. همچنین با بیشینه‌کردن لگاریتم شباهت نسبت به پارامترهای میانگین  $\mathbf{x}$  ها ( $\boldsymbol{\mu}_{xi}$ )، میانگین  $\mathbf{y}$  ها ( $\boldsymbol{\mu}_{yi}$ )،  $\delta_{xi}^2$  و  $\delta_{yi}^2$  مقادیر جدید برای این پارامترها به دست می‌آید. به علاوه پارامترهای مجھول هر مؤلفه را می‌توان با تجزیه مقدار ویژه ماتریس کواریانس محلی داده برای آن مؤلفه به دست آورد. ماتریس کواریانس محلی برای هر مدل به صورت زیر قابل تعریف است:

$$\mathbf{S}_i = \frac{1}{N} \sum_{n=1}^N \boldsymbol{\Phi}_i \frac{1}{2} \begin{pmatrix} \mathbf{x}_n - \boldsymbol{\mu}_{xi} \\ \mathbf{y}_n - \boldsymbol{\mu}_{yi} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n - \boldsymbol{\mu}_{xi} \\ \mathbf{y}_n - \boldsymbol{\mu}_{yi} \end{pmatrix}^T \quad (17)$$

$$\boldsymbol{\Phi}_i \frac{1}{2} = \begin{pmatrix} \frac{1}{\delta_{xi}^2} \mathbf{S}_{xi} & \frac{1}{\delta_{xy}^2} \mathbf{S}_{xyi} \\ \frac{1}{\delta_{xy}^2} \mathbf{S}_{yxi} & \frac{1}{\delta_{yi}^2} \mathbf{S}_{yi} \end{pmatrix} \quad (18)$$

که در این رابطه  $\mathbf{S}_{xi}$  به صورت زیر تعریف می‌شود:

$$\mathbf{S}_{xi} = \frac{1}{\pi_i N} \sum_{n=1}^N \mathbf{R}_{ni} (\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})^T \quad (19)$$

ماتریس‌های  $\mathbf{S}_{xyi}$ ،  $\mathbf{S}_{yxi}$  و  $\mathbf{S}_{yyi}$  نیز به صورت مشابه با رابطه بالا محاسبه می‌شوند.

با به دست آوردن مقادیر ویژه و بردارهای ویژه ماتریس کواریانس بالا می‌توان  $\mathbf{W}_{xi}$  و  $\mathbf{W}_{yi}$  برای هر مدل را با استفاده از روابط زیر به دست آورد:

$$\begin{aligned}\mathbf{W}_{xi} &= \delta_{xi} \mathbf{U}_{xi} (\mathbf{\Lambda}_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R} \\ \mathbf{W}_{yi} &= \delta_{yi} \mathbf{U}_{yi} (\mathbf{\Lambda}_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R}\end{aligned}\quad (20)$$

که  $\mathbf{U}_{xi}$  و  $\mathbf{\Lambda}_K = \text{diag}(\lambda_1, \dots, \lambda_K)$  شامل  $M$  سطر نخست از ماتریس  $[\mathbf{u}_{1i}, \dots, \mathbf{u}_{Ki}]$  و  $\mathbf{U}_{yi}$ ،  $\mathbf{\Lambda}_K$  سطر باقیمانده از آن است. ماتریس  $\mathbf{R}$  یک ماتریس معتمد دلخواه  $K \times K$  است.

نگاشت  $\mathbf{z}^*$  برای ورودی جدید  $\mathbf{x}^*$  در فضای کاهش بعد یافته به صورت زیر خواهد بود:

$$\mathbf{z}^* = \frac{1}{\delta_{xi}} \mathbf{R}^T (\mathbf{\Lambda}_K - \mathbf{I})^{-\frac{1}{2}} [\mathbf{U}_{xi}^T \mathbf{U}_{xi} + (\mathbf{\Lambda}_K - \mathbf{I})^{-1}]^{-1} \mathbf{U}_{xi}^T (\mathbf{x}^* - \boldsymbol{\mu}_{xi}) \quad (21)$$

فصل نهم



دیدگاه مدل جستجو، همه نامزدهایی که در  $M_p$  نیستند از دامنه جستجو حذف می‌شوند.

بدین ترتیب کاهش بعد باعث ازبین‌رفتن اطلاعات می‌شود، و هیچ تضمینی وجود ندارد که پارامترهای به دست آمده در فضای کاهش بعدياً فتله با پارامترهای بهینه در فضای اصلی یکسان باشند؛ لذا در روش جرمیه نگاشت پیشنهاد شده است که جستجوی پارامترها در فضای اصلی انجام شود، و فاصله آن از زیرفضای به دست آمده به عنوان جرمیه در نظر گرفته شود (زانگ و اشنایدر، ۲۰۱۰). با این روش مدل پیش‌گو به صورت زیر در خواهد آمد:

$$\operatorname{argmin}_{\tilde{\mathbf{w}}} \sum_{i=1}^N L(y_i, \tilde{\mathbf{w}}^T \mathbf{x}_i + \mathbf{v}^T (\mathbf{P} \mathbf{x}_i) + b) + \lambda J(\tilde{\mathbf{w}}) \quad (30)$$

که  $\tilde{\mathbf{w}} = \mathbf{w} - \mathbf{P}^T \mathbf{v}$  و  $J(\tilde{\mathbf{w}})$  تابع جرمیه است و می‌تواند به صورت  $\| \cdot \|_2^2$  یا  $\| \cdot \|_1$  در نظر گرفته شود. متغیر  $\lambda$  نیز پارامتر تنظیم است. اگر مدل پیش‌گویی مورد نظر و عمل‌گر کاهش بعد طوری طراحی شوند که روی فضای هسته کار کنند، مدل پیش‌گو با جرمیه نگاشت به صورت زیر خواهد بود:

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^p, b} \sum_{i=1}^N L(y_i, \mathbf{w}^T \Phi(\mathbf{x}_i) + b) + \min_{\mathbf{v} \in \mathbb{R}^d} \lambda J(\mathbf{w} - \mathbf{P}^T \mathbf{v}) \quad (31)$$

که  $\Phi(\mathbf{x}_i)$  بازنمایی نمونه  $\mathbf{x}_i$  در فضای هسته است. از آن جا که مسئله، دسته‌بندی است، در رابطه بالا از SVM با متغیرهای سستی  $\xi_i^1$  به عنوان تابع  $L$  استفاده می‌شود. به این ترتیب مسئله برنامه‌ریزی درجه دوی زیر را داریم:

$$\begin{aligned} & \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^d, b, \xi_i \in \mathbb{R}, i=1}^N \frac{1}{2} \left\| \mathbf{w} - \mathbf{P}^T \mathbf{v} \right\|_2^2 + C \sum_{i=1}^n \xi_i \\ & \text{s.t. } y_i (\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \forall i \\ & \quad \xi_i \geq 0 \quad \forall i \end{aligned} \quad (32)$$

که  $c \propto \frac{1}{\lambda}$  به جای  $\lambda$  در رابطه (۳۱) قرار گرفته است. با فرازدادن  $\tilde{\mathbf{w}} = \mathbf{w} - \mathbf{P}^T \mathbf{v}$  داریم [۱۶]:

$$\begin{aligned} & \operatorname{argmin}_{\tilde{\mathbf{w}} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^d, b, \xi_i \in \mathbb{R}, i=1}^N \frac{1}{2} \left\| \tilde{\mathbf{w}} \right\|_2^2 + C \sum_{i=1}^n \xi_i \\ & \text{s.t. } y_i (\tilde{\mathbf{w}}^T \Phi(\mathbf{x}_i) + \mathbf{v}^T \mathbf{P} \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \forall i \\ & \quad \xi_i \geq 0 \quad \forall i \end{aligned} \quad (33)$$

<sup>۱</sup> Slack Variables

<sup>۲</sup> Quadratic Programming

$$\begin{aligned} & \langle \mathbf{z}_{ni} \rangle = \mathbf{A}_i^{-1} \left[ \frac{1}{2} \frac{\mathbf{W}_{xi}^T (\mathbf{x}_n - \boldsymbol{\mu}_{xi})}{\delta_{xi}} + \frac{1}{2} \frac{\mathbf{W}_{yi}^T (\mathbf{y}_n - \boldsymbol{\mu}_{yi})}{\delta_{yi}} \right] \\ & \langle \mathbf{z}_{ni} \mathbf{z}_{ni}^T \rangle = \mathbf{A}_i^{-1} + \langle \mathbf{z}_{ni} \rangle \langle \mathbf{z}_{ni} \rangle^T \end{aligned} \quad (26)$$

در شکل (۲) کارنمای روش پیشنهادی نمایش داده است.

### ۲-۳- کاهش بعد محلی خطی بدون اتلاف با استفاده از جرمیه نگاشت

همان‌طور که پیش‌تر گفته شد، در این مقاله روشی برای شناسایی چهره با استفاده از کاهش بعد بدون اتلاف ارائه می‌شود. یادگیری یک مدل پیش‌بینی خطی با پارامترهای  $(\mathbf{w}, b)$  در یک فضای  $p$  بعدی با  $N$  نمونه آزمایشی به صورت زیر قابل تعریف است:

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^p} \sum_{i=1}^n L(y_i, \mathbf{w}^T \mathbf{x}_i + b) \quad (27)$$

که  $\mathbf{w} \in \mathbb{R}^p$  بردار پارامتر و  $b$  پارامتر مدل پیش‌گو را بازنمایی می‌کند. متغیرهای  $\mathbf{x}_i$  و  $y_i$  به ترتیب بردار ویژگی  $p$  بعدی و خروجی مربوط به نمونه  $i$ ام را مشخص می‌کنند. تابع  $L$  یک تابع اتلاف است، و به مدل پیش‌گو بستگی دارد که می‌تواند به صورت مربعات خطأ در نظر گرفته شود. اگر  $\mathbf{P}$  را یک ماتریس با ابعاد  $d \times p$  در نظر بگیریم، که یک کاهش بعد خطی به  $d$  بعد بر روی فضای ورودی  $p$  بعدی انجام می‌دهد، بازنمایی کاهش بعدياً فتله ورودی  $\mathbf{x}$  به صورت  $\mathbf{P} \mathbf{x}$  خواهد بود. با اعمال کاهش بعد خطی مدل پیش‌گو در فضای کاهش بعد بعدياً فتله به صورت زیر تعریف می‌شود:

$$\operatorname{argmin}_{\mathbf{v} \in \mathbb{R}^d, b} \sum_{i=1}^N L(y_i, \mathbf{v}^T (\mathbf{P} \mathbf{x}_i) + b) \quad (28)$$

که  $\mathbf{v} \in \mathbb{R}^d$  بردار پارامتر یادگرفته شده در فضای کاهش بعدياً فتله است.

از مقایسه رابطه (۲۷) با (۲۸) ارتباط بین کاهش بعد فضای ویژگی و محدودیت روی فضای پارامتر را می‌بینیم. استفاده از ماتریس کاهش بعد  $\mathbf{P}$  در فضای ورودی، موجب می‌شود که بردار پارامتر  $\mathbf{w}$  با بعد  $p$  بر زیرفضای زیر واقع شود:

$$M_p = \{ \mathbf{w} \in \mathbb{R}^p \mid \mathbf{w} = \mathbf{P}^T \mathbf{v}, \exists \mathbf{v} \in \mathbb{R}^d \} \quad (29)$$

بنابراین اعمال عمل‌گر کاهش بعد خطی  $\mathbf{P}$  در فضای ورودی معادل با محدود کردن مدل جستجو به یک زیرفضای پارامتر  $M_p$  است، که در رابطه (۲۹) تعریف شده است. از

است، استفاده گردید. در این هنجارسازی، تصاویر بر اساس فاصله بین دو چشم تنظیم شده و سپس تصویر چهره از عکس استخراج می‌شود. پایگاه داده ORL شامل تصاویری از چهل نفر است. از هر نفر ده تصویر متفاوت وجود دارد. تصاویر این پایگاه داده، تغییراتی از قبیل شدت روشنایی، حالت‌های چهره (چشمان باز و بسته، و لبخند)، تصاویر با عینک و بدون عینک، مقیاس و فاصله و چرشش زاویه سر را شامل می‌شود. در آزمایش بر روی این پایگاه داده‌ها ابتدا هر تصویر به  $32 \times 32$  پیکسل تغییر اندازه داده می‌شود.

در آزمایش‌ها از روش استخراج ویژگی گابور<sup>۳</sup> برای بهدست آوردن بردار ویژگی مربوط به هر تصویر استفاده شده است. فیلترهای گابور ابزاری برای استخراج ویژگی هستند. این توابع دوبعدی، لبۀ اشکال و همچنین گودی‌ها و برآمدگی‌های تصویر را تقویت می‌کنند. برای افزایش تمایز چشم‌ها، دهان و بینی که به عنوان اجزای مهم و اساسی چهره مطرح هستند، فیلترهای گابور می‌توانند مفید باشند (حقیقت و همکاران ۲۰۱۳). طول بردار ویژگی استخراج شده با این روش ۲۵۶۰ است.

در آزمایش‌ها روش‌های زیر با هم مقایسه شده‌اند: (۱) روش ماشین‌بردار پشتیبان (SVM) در فضای ویژگی کاهش‌بعدیافته با روش تحلیل مؤلفه اصلی هسته (KPCA) برای کاهش بعد، که به اختصار آن را SVM-KPCA می‌نامیم. این روش ابتدا KPCA را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کند. (۲) روش ماشین‌بردار پشتیبان در فضای ویژگی کاهش‌بعدیافته با روش تحلیل مؤلفه اصلی احتمالاتی که آن را به اختصار SVM-PPCAMM می‌نامیم. این روش ابتدا PPCAMM را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کند. (۳) روش ماشین‌بردار پشتیبان در فضای ویژگی کاهش‌بعدیافته با روش پیشنهادی مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظرارت که آن را به اختصار SVM-SPPCAMM می‌نامیم. در این روش ابتدا PPCAMM را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کنیم. (۴) روش با Proj-KPCA در درنظرگرفتن جریمه نگاشت که به اختصار آن را Proj-KPCA می‌نامیم. (۵) مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی با درنظرگرفتن جریمه نگاشت، که آن را به اختصار Proj-PPCAMM می‌نامیم. (۶) به کارگیری روش پیشنهادی

تا اینجا کاهش بعد با درنظرگرفتن جریمه نگاشت برای روش‌های کاهش بعد خطی و غیرخطی مبتنی بر هسته بیان شد. آنچه در این مقاله مورد بررسی قرار می‌گیرد، کاهش بعد محلی خطی با درنظرگرفتن جریمه نگاشت است. در روابط (۳۰) و (۳۳) عمل گر کاهش بعد  $\mathbf{P}$  تنها از طریق جمله  $\mathbf{P}\mathbf{x}_i$  ( $\mathbf{P}$  در رابطه  $(\mathbf{x}_i)$ ) و  $\mathbf{P}\Phi(\mathbf{x}_i)$  ( $\Phi$  در رابطه  $(\mathbf{x}_i)$ ) ظاهر شده است. می‌توان با یک ترفند ساده این دو جمله را با یک تابع کاهش بعد عمومی مانند  $\Psi(\mathbf{x}_i)$  جایگزین کرد [۱۶]. در این مقاله یک خمینه<sup>۱</sup> زیربنایی محلی خطی با استفاده از روش ارائه شده برای مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظرارت (SPPCAMM) از نمونه داده‌ها بهدست می‌آید، که همان تابع کاهش بعد  $(\mathbf{x}_i)\Psi$  خواهد بود. با جایگذاری این تابع به جای  $\mathbf{P}$  در رابطه (۳۳) مسئله بهینه‌سازی زیر را خواهیم داشت:

$$\begin{aligned} \operatorname{argmin}_{\tilde{\mathbf{w}} \in \mathbb{R}^p, v \in \mathbb{R}^d, b, \xi_i | i=1 \dots n} & \frac{1}{2} \left\| \tilde{\mathbf{w}} \right\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } & \mathbf{y}_i \left( \tilde{\mathbf{w}}^T \Phi(\mathbf{x}_i) + \mathbf{v}^T \Psi(\mathbf{x}_i) + b \right) \geq 1 - \xi_i, \quad \forall i \\ & \xi_i \geq 0 \quad \forall i \end{aligned} \quad (34)$$

درواقع در این روش به جای نگاشت داده بر خمینه زیربنایی و جستجوی پارامترهای مدل پیش‌گو در آن که موجب از دست‌رفتن اطلاعات می‌شود، همان‌طور که در روابط بالا بیان شد، جستجو را در فضای اصلی انجام می‌دهیم و فاصله پارامترهای بهدست‌آمده در فضای اصلی از پارامترهای بهدست‌آمده در خمینه محلی خطی حاصل از SPPCAMM را به عنوان یک جریمه نگاشت در یادگیری مدل پیش‌گو مورد استفاده قرار می‌دهیم.

## ۴- نتایج پیاده‌سازی

در این بخش به بررسی کارایی روش پیشنهادی پرداخته و آن را با چند روش کاهش بعد دیگر مورد مقایسه قرار می‌دهیم. برای این منظور مسئله شناسایی چهره را در نظر گرفته‌ایم، و آزمایش‌ها را بر روی پایگاه داده‌های چهره ORL و Yale انجام داده‌ایم. پایگاه داده Yale شامل ۱۶۵ تصویر سیاه‌سفید از پانزده نفر است. برای هر شخص یازده تصویر با حالت‌های مختلف چهره و شرایط نوری متفاوت وجود دارد. در این مقاله از تصاویر این پایگاه داده که بر روی سایت دانشگاه MIT قرار دارد<sup>۲</sup> و شامل تصاویر هنجارسازی شده

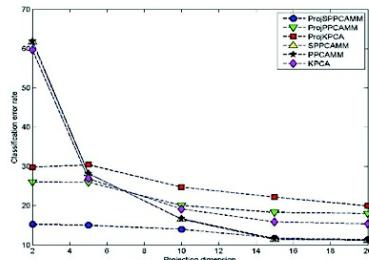
<sup>1</sup> Manifold

<sup>2</sup> <http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets/>

<sup>3</sup> Gabor



به تعداد کل چهره‌ها برای ارزیابی گزارش شده است. آزمایش‌ها برای سی بار اجرای تصادفی بر روی پایگاه داده Yale با مجموعه آموزشی شامل هفت تصویر از هر شخص و کاهش بعدی‌بعد ده انجام شده است.



(شکل ۱) نمودار نرخ خطای دسته بنده (به درصد) در مقابل تعداد ویژگی‌های استخراج شده

در این جدول همچنین دو معیار ارزیابی دقت<sup>۲</sup> و فراخوانی<sup>۳</sup> برای یکبار اجرای تصادفی با همان تنظیمات قبلی گزارش شده است. این مقادیر از میانگین دقت و فراخوانی دسته‌های مختلف مسأله روی نمونه‌های آزمایشی، که در ادامه تعریف شده‌اند، محاسبه می‌شوند. دقت دسته نام نسبت تعداد نمونه‌هایی است که الگوریتم به درستی آن‌ها را به دسته نام منسوب کرده است، به تعداد کل نمونه‌هایی که به درستی متعلق به دسته نام هستند.

همچنین در جدول (۶) روش‌های مورد مقایسه از نظر زمان اجرا مورد بررسی قرار گرفته‌اند. زمان‌های اجرای به دست آمده برای کاهش بعدی‌بعد ده بر روی پایگاه داده Yale و مجموعه آموزشی شامل سه تصویر از هر شخص بوده و بر حسب ثانیه گزارش شده‌اند.

از آن جا که روش پیشنهادی یک روش بانظارت است، با روش‌های کاهش بعد بانظارت Yale نیز مقایسه شده است. برای این منظور از پایگاه داده Yale استفاده شده، و کاهش بعدی‌بعد ده بر روی نمونه داده‌ها انجام گرفته است. برای روش‌های مورد بررسی از مجموعه داده‌های آموزشی و آزمایشی یکسان استفاده شده است. نتایج این آزمایش‌ها در جدول شماره (۷) آمده است.

SPPCAMM با جریمه نگاشت، که آن را به اختصار -Proj-SPPCAMM می‌نامیم.

در جدول (۱) تا (۴) در همه روش‌ها نمونه داده‌ها به فضایی با بعد ده کاهش بعد داده شده‌اند. در آزمایش‌های انجام‌شده هر بار به صورت تصادفی تعدادی از نمونه‌ها برای آموزش در نظر گرفته شده و مابقی برای آزمایش استفاده می‌شوند. ما از مجموعه‌های آموزشی با ۳، ۵، و ۷ تصویر از هر شخص برای مقایسه روش‌ها استفاده کرده‌ایم. مقدار نخستین پارامترهای  $\delta_x^2$  و  $\delta_y^2$  در روش‌هایی که این پارامتر را دارند، برابر  $10^{-5}$  قرار داده شده است. هر آزمایش سی بار و به صورت مستقل تکرار شده است. در آزمایش‌هایی که در آن‌ها از روش استخراج ویژگی استفاده شده است، چون ویژگی‌های استخراج شده دارای مقادیر عددی کوچک هستند، برای عملکرد بهتر روش‌ها به خصوص روش‌های SPPCAMM و PPCAMM مقادیر با ضرب در یک مقدار عددی مناسب تغییر مقیاس داده شده‌اند.

جزیيات مربوط به پیاده‌سازی جریمه نگاشت به صورتی که در (زانگ و اشتایریدر، ۲۰۱۰) بیان شده، در نظر گرفته شده است. در استفاده از SVM با به کار گیری جریمه نگاشت از هسته چندجمله‌ای درجه دو استفاده شده است. همچنین پارامتر C در آزمایش‌ها برابر  $10^{-3}$  قرار داده شده است. در جدول (۱) و (۲) میانگین خطای دسته بنده برای هر روش بر روی پایگاه داده ORL گزارش شده است، و جدول (۳) و (۴) نتایج مربوط به پیاده‌سازی‌ها بر روی پایگاه داده Yale را نشان می‌دهند.

برای بررسی بیشتر روش‌های مورد مقایسه، به ازای تعداد ویژگی‌های استخراج شده مختلف، نه فقط ده ویژگی، آزمایش‌ها را تکرار کرده‌ایم. در آزمایش‌ها بر روی پایگاه داده Yale با مجموعه آموزشی شامل هفت تصویر از هر شخص، بعد داده به ۲، ۵، ۱۰، ۱۵، و ۲۰ کاهش داده شده، و متوسط نرخ خطای دسته بنده در مقابل تعداد ویژگی‌های استخراج شده برای سی بار اجرای تصادفی در قالب نمودار شکل (۱) رسم شده است. در این آزمایش‌ها خطای به دست آمده برای روش SPPCAMM در بعدهای مختلف به خطاهای روش PPCAMM بسیار نزدیک است. از این رونمودار خطای این روش‌ها در این شکل بر روی هم قرار گرفته است.

در جدول (۵) معیار نرخ شناسایی درست<sup>۱</sup> به معنای نسبت تعداد چهره‌هایی که درست تشخیص داده می‌شوند،

<sup>1</sup> Recognition rate

<sup>2</sup> Precision  
<sup>3</sup> Recall

## ۵- تحلیل نتایج

همان طور که از نتایج به دست آمده از آزمایش‌ها مشاهده می‌شود، روش پیشنهادی با درنظر گرفتن جرمیمه نگاشت (Proj-SPPCAMM) در مقایسه با سایر روش‌های کاهاش بعد آزمایش شده، در تمام آزمایش‌ها خطای دسته‌بندی کمتری دارد. با مقایسه روش PPCAMM با درنظر گرفتن Proj-SPPCAMM جرمیمه نگاشت و مدل باظارت آن یعنی Proj-SPPCAMM برتری روش PPCAMM به طور کامل مشهود است. این به بدان معناست که با استفاده از اطلاعات خروجی و اعمال جرمیمه نگاشت، قادر به یادگیری بهتر مدل در مسائل باظارت هستیم. به علاوه همان طور که از نتایج آزمایش‌ها مشاهده می‌شود، روش PPCAMM در مقایسه با KPCA در بیش‌تر آزمایش‌ها خطای کمتری دارد. این امر می‌تواند نشان‌دهنده مناسب تربووند روش‌های کاهاش بعد محلی خطی نسبت به روش‌های کاهاش بعد غیر خطی سراسری در ارائه مجموعه داده‌ها (به دلیل انعطاف‌پذیری بیش‌تر و ساده تربووند این مدل‌ها) باشد. از آن‌جا که مدل باظارت این روش، یعنی SPPCAMM که در این مقاله ارائه شده است، هم می‌تواند از هم دارای خاصیت محلی خطی است، و هم می‌تواند از اطلاعات مربوط به خروجی در مسائل پیش‌بینی و دسته‌بندی بهره گیرد، برتری این مدل نسبت به دیگر روش‌های کاهاش بعد خطی و غیرخطی موجه بوده، و این ویژگی‌ها امکان استفاده از آن در مسائل پیچیده و باظارت را فراهم می‌کند.

مقایسه روش پیشنهادی با روش LDA در جدول (۷) نیز برتری روش پیشنهادی به خصوص برای تعداد داده‌های آموزشی کم را نشان می‌دهد؛ مقایسه نتایج روش SPPCA و SPPCAMM برتری نسبی روش SPPCAMM نشان می‌دهد؛ اما همان‌طور که از جدول (۷) مشاهده می‌شود، نتایج بسیار به هم نزدیک هستند. این امر را می‌توان به دلیل کوچک‌بودن مجموعه آموزشی و یا در حالت کلی کوچک‌بودن مجموعه نمونه داده‌ها دانست. چون برای یک مجموعه داده کوچک چند مؤلفه در نظر گرفته می‌شود؛ در حالی که یک مؤلفه هم می‌تواند بازنمایی داده را به خوبی انجام دهد. اگر مجموعه داده بزرگ باشد استفاده از روش SPPCAMM و درنظر گرفتن چند مؤلفه برای بازنمایی داده کارایی را به نحو قابل توجهی می‌تواند بهبود بخشد.

همان‌طور که از نتایج به دست آمده از آزمایش‌ها مشاهده می‌شود، با افزایش تعداد داده‌های آموزشی متوجه خطای دسته‌بندی به طور تقریبی جدا از هم در فضای

می‌یابد، که این ویژگی برای روش پیشنهادی نیز برقرار است. همچنین با توجه به نمودارهای شکل (۱) که خطای دسته‌بندی در مقابل تغییرات بعد را نشان می‌دهد، مشاهده می‌شود که با افزایش بعد، خطای دسته‌بندی کاهاش می‌یابد. بررسی زمان اجرای روش‌های آزمایش شده در جدول (۶) نشان می‌دهد که زمان اجرای روش پیشنهادی- Proj-SPPCAMM نسبت به سایر روش‌ها بیشتر است. بعد از این روش و روش SPPCAMM زمان اجرای روش- Proj- PPCAMM و PPCAMM زیاد است. در این روش‌ها قسمت بیش‌تر زمان اجرا مربوط به آموزش است؛ که دلیل آن استفاده از روش تکراری الگوریتم EM در آن‌ها است. زمان بالای آموزش را می‌توان ضعف روش پیشنهادی دانست؛ اما دقیق به دست آمده از این روش به خصوص در بعدهای پایین قابل توجه است.

در روش پیشنهادی در بخش ۱-۳ روابط با فرض گاسی بودن توزیع داده به دست آمده است؛ لذا ممکن است این سؤال به ذهن آید که اگر توزیع داده غیر گاسی باشد آیا این روش همچنان عملکرد قابل قبولی خواهد داشت؟ همان‌طور که می‌دانیم در مسائل دنیای واقعی فرض گاسی بودن داده همیشه برقرار نیست. در مسائل بررسی شده در این مقاله، یعنی شناسایی چهره بر روی پایگاه داده Yale ORL، نیز به الزام فرض گاسی بودن داده برقرار نیست. این در حالی است که نتایج آزمایش‌ها نشان می‌دهند روش پیشنهادی برای این مجموعه داده‌ها عملکرد خوبی داشته است؛ لذا مشاهده می‌شود که روش پیشنهادی برای داده غیر گاسی هم عملکرد قابل قبولی دارد.

به منظور کسب بینش شهودی در این زمینه سه توزیع یکنواخت در فضای سه‌بعدی را که دامنه آنها سه مکعب نزدیک به یکدیگر هستند، در نظر گرفته و از هر یکصد نمونه به تصادف برداشته‌ایم. بدین ترتیب چنان که در شکل (۲) نشان داده شده است، سیصد نمونه داده با بعد سه به دست آمده است. برای نمونه‌های تولیدشده از هر توزیع یک برچسب دسته در نظر گرفته‌ایم. بدین ترتیب نمونه‌ها از سه دسته متفاوت در نظر گرفته می‌شوند؛ سپس با استفاده از روش پیشنهادی SPPCAMM با سه مؤلفه، نمونه داده‌ها را به فضای با بعد دو نگاشت کرده‌ایم. شکل (۳) نمونه‌های کاهاش بعدیافته بدین ترتیب را نشان می‌دهد. همان‌طور که در شکل (۳) مشاهده می‌شود، نمونه داده‌ها به صورت سه دسته‌تۀ به طور تقریبی جدا از هم در فضای

## ۶- جمع‌بندی

پیشنهادهایی به شرح زیر برای ارتقای روش پیشنهادی در این مقاله مطرح هستند. همان‌طور که بیان شد، در این مقاله از جرمیه نگاشت با دسته‌بند SVM استفاده شده است. به کارگیری ایده جرمیه نگاشت با استفاده از سایر روش‌های دسته‌بندی، به عنوان یک موضوع پژوهشی قابل بررسی است. همچنین در روش کاهش بعد بدون اتلاف به کاررفته در این مقاله از هسته چندجمله‌ای استفاده شده است. استفاده از هسته‌های دیگر و نیز یادگیری چند‌هسته‌ای در این مورد می‌تواند مفید باشد.

(جدول-۱): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده ORL برای ۳۰ بار اجرای تصادفی

ORL Dataset - Classification error rate (in %) $\pm$ standard deviation ( $p$ -value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	18.39 $\pm$ 3.27	18.01 $\pm$ 3.40	16.83 $\pm$ 2.88	15.50 $\pm$ 2.54	16.63 $\pm$ 3.47	<b>10.37 <math>\pm</math> 2.37</b>
5	9.08 $\pm$ 2.64	8.85 $\pm$ 2.27	8.37 $\pm$ 2.69	7.77 $\pm$ 2.01	9.71 $\pm$ 2.40	<b>4.25 <math>\pm</math> 1.92</b>
7	5.61 $\pm$ 1.61	5.31 $\pm$ 2.42	5.58 $\pm$ 3.85	4.22 $\pm$ 1.69	4.86 $\pm$ 2.10	<b>2.14 <math>\pm</math> 1.38</b>

(جدول-۲): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده ORL استخراج ویژگی شده به روش گابور برای ۳۰ بار اجرای تصادفی

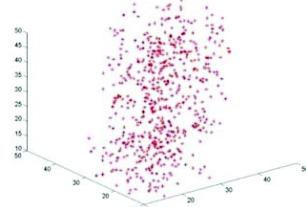
ORL Dataset – Gabor Feature Extraction- Classification error rate (in %) $\pm$ standard deviation ( $p$ -value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	21.61 $\pm$ 2.56	17.99 $\pm$ 3.91	17.99 $\pm$ 3.91	18.07 $\pm$ 3.11	14.90 $\pm$ 3.16	<b>7.23 <math>\pm</math> 1.98</b>
5	11.00 $\pm$ 2.23	9.91 $\pm$ 2.21	9.91 $\pm$ 2.21	9.32 $\pm$ 1.99	8.07 $\pm$ 1.76	<b>1.96 <math>\pm</math> 0.84</b>
7	7.22 $\pm$ 1.54	5.95 $\pm$ 1.76	5.95 $\pm$ 1.76	6.31 $\pm$ 1.77	5.92 $\pm$ 1.70	<b>1.06 <math>\pm</math> 0.90</b>

(جدول-۳): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده Yale برای ۳۰ بار اجرای تصادفی

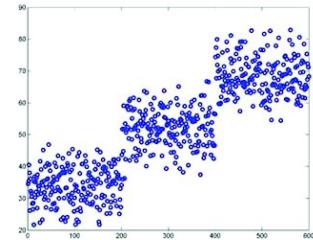
Yale Dataset - Classification error rate (in %) $\pm$ standard deviation ( $p$ -value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	25.91 $\pm$ 3.85	26.52 $\pm$ 3.11	26.52 $\pm$ 3.11	25.96 $\pm$ 3.46	26.17 $\pm$ 2.63	<b>19.50 <math>\pm</math> 2.67</b>
5	22.06 $\pm$ 4.10	21.62 $\pm$ 3.38	21.62 $\pm$ 3.38	21.97 $\pm$ 4.16	21.70 $\pm$ 2.85	<b>16.27 <math>\pm</math> 3.55</b>
7	20.11 $\pm$ 3.63	18.62 $\pm$ 3.97	18.01 $\pm$ 4.27	21.39 $\pm$ 4.91	20.84 $\pm$ 3.55	<b>15.67 <math>\pm</math> 3.55</b>

کاهش بعدیافته ظاهر شده‌اند، که بیان گر کارآی مناسب روش پیشنهادی برای داده با توزیع غیر گاسی است.

در این مقاله ابتدا مدل با نظرارت روش ترکیبی تحلیل مؤلفه اصلی احتمالاتی (SPPCAMM) ارائه شده است. این مدل توسعه‌ای از مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی است، که اطلاعات مربوط به برچسب داده را در نگاشت یادگیری شده منظور می‌کند. شکل (۴) روندمنای این مدل پیشنهادی را نشان می‌دهد؛ سپس این رویکرد با نظرارت کاهش بعد محلی خطی SPPCAMM در چارچوبی به نام جرمیه نگاشت با استفاده از دسته‌بند SVM به کار گرفته شده است. جرمیه نگاشت، این امکان را فراهم می‌آورد که از کاهش بعد برای بهبود دقت مدل پیش‌گو بدون خطر از دستدادن اطلاعات مرتبط استفاده کنیم. نتایج آزمایش‌ها نشان‌دهنده برتری روش پیشنهادی (Proj-SPPCAMM) با درنظرگرفتن جرمیه نگاشت (Proj-SPPCAMM) در مقایسه با سایر روش‌های کاهش بعد بررسی شده در زمینه دقت دسته‌بندی است.



(شکل-۲): نمونه داده‌های تصادفی به دست آمده از سه توزیع یکنواخت نزدیک به هم



(شکل-۳): نسخه کاهش بعد یافته نمونه داده‌های تصادفی شکل (۲) با روش SPPCAMM

(جدول-۴): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده Yale، استخراج ویزگی شده به روش گابور برای ۳۰ بار اجرای تصادفی

Yale Dataset – Gabor Feature Extraction- Classification error rate (in %) $\pm$ standard deviation ( <i>p</i> -value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	16.78 $\pm$ 2.89	14.17 $\pm$ 3.52	13.63 $\pm$ 3.11	14.97 $\pm$ 2.49	14.03 $\pm$ 2.98	<b>10.46<math>\pm</math>2.65</b>
5	13.11 $\pm$ 2.80	11.87 $\pm$ 3.13	11.85 $\pm$ 3.07	12.78 $\pm$ 2.69	12.07 $\pm$ 1.92	<b>8.11<math>\pm</math>2.80</b>
7	11.83 $\pm$ 3.51	10.97 $\pm$ 4.18	10.97 $\pm$ 4.18	11.17 $\pm$ 4.38	11.56 $\pm$ 3.06	<b>7.74<math>\pm</math>3.41</b>

(جدول - ۵): بررسی معیارهای ارزیابی نرخ شناسایی درست، دقت، و فراخوانی بر روی پایگاه داده Yale و برای مجموعه آموزشی با ۷ تصویر از هر شخص

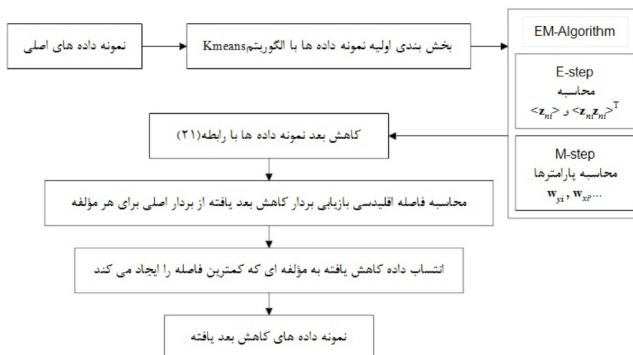
	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
Recognition rate	0.81	0.83	0.84	0.75	0.82	0.85
Precision	0.90	0.93	0.94	0.93	0.94	0.93
Recall	0.88	0.90	0.85	0.92	0.92	0.90

(جدول-۶): بررسی زمان آموزش و آزمایش روش‌های مورد مقایسه بر حسب ثانیه

	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
زمان آزمایش	0.0960	0.1968	0.4242	3.8218	3.6171	4.8276
زمان آموزش	0.8156	43.5106	56.1545	3.0556	46.4498	72.4550

(جدول-٧): متوسط خطای دسته‌بندی روش‌های بانظار مختلط بر روی پایگاه داده Yale ۳۰ بار اجرای تصادفی

Yale Dataset - Classification error rate (in %) $\pm$ standard deviation ( <i>p</i> -value)						
Training faces(#)	LDA	SPPCA	SPPCAMM	Proj- LDA	Proj- SPPCA	Proj-SPPCAMM
3	32.32 $\pm$ 3.21	23.58 $\pm$ 3.91	23.61 $\pm$ 3.46	19.38 $\pm$ 3.01	18.83 $\pm$ 2.78	18.83 $\pm$ 2.78
5	31.89 $\pm$ 4.90	17.81 $\pm$ 3.45	17.78 $\pm$ 3.48	15.63 $\pm$ 2.87	15.22 $\pm$ 2.66	15.22 $\pm$ 2.66
7	27.44 $\pm$ 3.48	17.56 $\pm$ 4.27	17.56 $\pm$ 4.27	14.78 $\pm$ 3.62	15.17 $\pm$ 3.73	15.07 $\pm$ 3.57



(شکل - ۴): کارنمای روش پیشنهادی SPPCAMM

Control, 2006. ICICIC'06. First International Conference on, 2006, pp. 345-348.

C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

F. Han, T. Zhao, and H. Liu, "Coda: High dimensional copula discriminant analysis," *Journal of Machine Learning Research*, vol. 14, pp. 629-671, 2013.

٧- مراجع

C. Chen, J. Zhang, and R. Fleischer, "Distance approximating dimension reduction of Riemannian manifolds," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, pp. 208-217, 2010.

C.-G. Li and J. Guo, "Supervised isomap with explicit mapping," in *Innovative Computing, Information and*

Marginal Discriminant Embedding," Procedia Engineering, vol. 29, pp. 494-498, 2012.



سمیه احمدخانی کارشناسی خود را در رشته مهندسی کامپیوتر گرایش سخت افزار در سال ۱۳۸۹ از دانشگاه صنعتی همدان اخذ کرد. ایشان هم‌اکنون دانشجوی کارشناسی ارشد

رشته مهندسی کامپیوتر گرایش هوش مصنوعی در دانشگاه اصفهان است. زمینه‌های مورد علاقه ایشان یادگیری ماشین،

شناسایی الگو و پردازش تصویر می‌باشد.  
نشانی رایانامه ایشان عبارت است از:

s.ahmadkhani@eng.ui.ac.ir



پیمان ادبی مدرک کارشناسی خود را در رشته مهندسی کامپیوتر گرایش سخت افزار در سال ۱۳۷۶ از دانشگاه صنعتی اصفهان، و مدرک کارشناسی ارشد و دکترای خود را در رشته

مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه صنعتی امیرکبیر تهران به ترتیب در سال‌های ۱۳۸۰ و ۱۳۸۸ دریافت کرد. ایشان از سال ۱۳۸۹ تاکنون استادیار دانشکده مهندسی کامپیوتر دانشگاه اصفهان هستند. زمینه‌های پژوهشی مورد علاقه ایشان در حال حاضر شامل یادگیری

ماشین و شناسایی الگو، هوش محاسباتی و محاسبات نرم و بینایی کامپیوتر است.

نشانی رایانامه ایشان عبارت است از:

adibi@eng.ui.ac.ir

J. Chen and Y. Liu, "Locally linear embedding: a survey," Artificial Intelligence Review, vol. 36, pp. 29-48, 2011.

K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 2, pp. 559-572, 1901.

L. Van der Maaten, E. Postma, and H. Van Den Herik, "Dimensionality reduction: A comparative review," Journal of Machine Learning Research, vol. 10, pp. 1-41, 2009.

M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," Neural computation, vol. 11, pp. 443-482, 1999.

M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, pp. 611-622, 1999.

<http://vismod.media.mit.edu/vismod/classes/mas622-0/datasets/>

M. Haghigat, S. Zonouz, and M. Abdel-Mottaleb, "Identification Using Encrypted Biometrics," in Computer Analysis of Images and Patterns, 2013, pp. 440-448.

M. Zhu and A. M. Martinez, "Subclass discriminant analysis," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, pp. 1274-1286, 2006.

N. Kalealis, V. Mezaris, I. Kompatsiaris, and T. Statthaki, "Mixture subclass discriminant analysis link to restricted Gaussian model and other generalizations," Neural Networks and Learning Systems, IEEE Transactions on, vol. 24, pp. 8-21, 2013.

R.A. Fisher, "The Statistical Utilization of Multiple Measurements," Annals of Eugenics, vol. 8, pp. 376-386, 1938.

S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, "Fisher Discriminant Analysis with Kernels," Proc. IEEE Signal Processing Soc. Workshop Neural Networks for Signal Processing IX, pp. 41-48, 1999.

S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, pp. 40-51, 2007.

S. Yu, K. Yu, V. Tresp, H.-P. Kriegel, and M. Wu, "Supervised probabilistic principal component analysis," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 464-473.

Y. Zhang and J. G. Schneider, "Projection penalties: Dimension reduction without loss," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 1223-1230.

Y.-D. Lan, H. Deng, and T. Chen, "Dimensionality Reduction Based on Neighborhood Preserving and

