به نام خدا

درس یادگیری‌ماشین

تمرین پنجم

حدیثه مصباح

۸۱۰۱۰۲۲۵۳

پاسخ ۱.

سوال یک:

$$J(D)=\frac{\tilde{\mu}_1-\tilde{\mu}_2}{\tilde{s}_1+\tilde{s}_2}\xrightarrow[S_1=S_2\rightarrow P S_1^T=P S_2]{\tilde{s}_1=\tilde{s}_2} J(D)=\frac{\tilde{\mu}_1-\tilde{\mu}_2}{2\tilde{s}^2}$$

$$\xrightarrow[\text{MLRatio}]{\text{اثبات}}\frac{P(\tilde{D}\mid\tilde{\theta}_1)}{P(\tilde{D}\mid\tilde{\theta}_2)}$$

حال چون $P(y\mid\tilde{\theta}_1)$ رو داریم نیاز داریم جایگذاری کنیم

$$\longrightarrow\frac{\frac{1}{\sqrt{2\pi\tilde{s}}}e^{-\frac{(D^Ta-\tilde{\mu}_1)^2}{2\tilde{s}^2}}}{\frac{1}{\sqrt{2\pi\tilde{s}}}e^{-\frac{(D^Ta-\tilde{\mu}_2)}{2\tilde{s}^2}}}=\exp\left(\frac{(D^Ta-\tilde{\mu}_2)^2-(D^Ta-\tilde{\mu}_1)^2}{2\tilde{s}^2}\right)$$

$$\xrightarrow{\text{اتحاد}}\frac{(\underbrace{(D^Ta-\tilde{\mu}_2)-(D^Ta-\tilde{\mu}_1)}_{\tilde{\mu}_1-\tilde{\mu}_2})^2+2(D^Ta-\tilde{\mu}_2)(D^Ta-\tilde{\mu}_1)}{2\tilde{s}^2}$$

$$\exp\left(\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{2\tilde{s}^2}\right)\times\underbrace{\exp\left(\frac{\cancel{2}(D^Ta-\tilde{\mu}_2)(D^Ta-\tilde{\mu}_1)}{\cancel{2}\tilde{s}^2}\right)}_{\alpha}$$

$$\alpha\exp\left(\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{2\tilde{s}^2}\right)\quad\checkmark$$

پس می‌توان از از log Likelihood ratio به $J(w)$ رسید.

سوال 2:

$$J_1(w) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \longrightarrow FDR \qquad y = w^T x$$

$$E(y) = E[w^T x] = w^T E[x] = \mu_{y_i} = w^T \mu_{x_i}$$

$$\sigma_{iy}^2 = E[(w^T x - \mu_{x_i})^2] = w^T E[(x - \mu_{x_i})(x - \mu_{x_i})^T] w$$

$$\rightarrow \frac{\partial J}{\partial w} \Rightarrow \quad \frac{\partial J_1}{\partial \mu_1} = \frac{2(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}, \quad \frac{\partial J_1}{\partial \mu_2} = \frac{-2(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2} \qquad \frac{\partial J_1}{\partial \sigma_1^2} = \frac{-(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)^2}$$

$$\frac{\partial J_1}{\partial \sigma_2^2} = \frac{-(\mu_1 - \mu_2)^2}{(\sigma^2 + \sigma^2)^2}$$

$$\frac{\partial J_1}{\partial w} = \frac{\partial J_1}{\partial \mu_1} \mu_{x_1} + \frac{\partial J}{\partial \mu_2} \mu_{x_2} + \left(2 \frac{\partial J}{\partial \sigma^2} \Sigma_1 + 2 \frac{\partial J}{\partial \sigma_1^2}\right) w^* = 0$$

$$w^* = \frac{\underbrace{\frac{2(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}}_{\frac{\partial J_1}{\partial \mu_1}} \mu_{x_1} - \underbrace{\frac{2(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}}_{\frac{\partial J}{\partial \mu_2}} \mu_{x_2}}{2\left(\underbrace{\frac{-(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)^2}}_{\frac{\partial J}{\partial \sigma_1^2}} \Sigma_1 - \underbrace{\frac{(\mu_1 - \mu_2)^2 \Sigma_2}{(\sigma_1^2 + \sigma_2^2)^2 \Sigma_1 - \frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)^2}}}_{\frac{\partial J}{\partial \sigma_2^2}} \Sigma_2\right)}$$

ساده سازی $\longrightarrow$ $$w^* \overset{☆}{=} \frac{(\sigma_1^2 + \sigma_2^2)^2 (\mu_1 - \mu_2)}{(\sigma_1^2 + \sigma_2^2)(\mu_1 - \mu_2)^2} \times \frac{\mu_{x_1} - \mu_{x_2}}{\Sigma_1 + \Sigma_2}$$

قسمت ☆ مانند K یک ضریب اسکالر است پس

$$J(kw) = J_1(w) \rightarrow J(kw) = \frac{(k\mu_1 - k\mu_2)^2}{(k\sigma_1)^2 + (k\sigma_2)^2} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} = J_1(w)$$

در نهایت با توجه به خط بالا از ضرب اسکالر با تصرف نظر می‌شود

$$w^* = \frac{(\mu_1 - \mu_2)}{(\Sigma_1 + \Sigma_2)} \longrightarrow w = (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)$$

در نتیجه می‌توان گفت $w^* \rightarrow \max J(w_1)$

پاسخ 3.

الف)

تجزیه به مؤلفه‌های اصلی (PCA)

مشکلات:

۱. حافظه: در PCA، محاسبه و ذخیره‌سازی ماتریس کوواریانس داده‌های با ابعاد بالا می‌تواند نیاز به مقادیر بسیار زیادی حافظه داشته باشد.

۲. محاسباتی: محاسبه مقادیر و بردارهای ویژه برای ماتریس‌های بزرگ می‌تواند بسیار زمان‌بر و محاسباتی باشد.

۳. دقت عددی: در ابعاد بالا، خطاهای عددی می‌تواند به تجمع یافته و باعث کاهش دقت نتایج شود.

راه حل‌ها:

۱. Randomized PCA: این روش با استفاده از تکنیک‌های تصادفی برای تقریب بردارهای ویژه و مقادیر ویژه، به کاهش زمان و حافظه مورد نیاز کمک می‌کند.

۲. Incremental PCA: برای داده‌های بسیار بزرگ که نمی‌توانند به طور کامل در حافظه جای بگیرند، Incremental PCA داده‌ها را به صورت قسمت‌های کوچکتر پردازش می‌کند و به تدریج مدل PCA را به‌روزرسانی می‌کند.

۳. کتابخانه‌های بهینه‌سازی شده: استفاده از کتابخانه‌هایی که عملیات‌های محاسباتی را بر روی سخت‌افزارهای مدرن (مانند GPU) بهینه‌سازی می‌کنند.

تجزیه خطی تمییزی (LDA)

مشکلات:

۱. محاسبه ماتریس‌های داخلی و بین کلاس: در داده‌های با ابعاد بالا، محاسبه و ذخیره این ماتریس‌ها می‌تواند بسیار حجیم و زمان‌بر باشد.

۲. تعداد کلاس‌ها نسبت به تعداد نمونه‌ها: در صورتی که تعداد کلاس‌ها زیاد باشد، ماتریس‌های محاسبه شده ممکن است ناسازگار باشند و منجر به مشکلات محاسباتی شوند.

راه حل‌ها:

۱. LDA مبتنی بر هیستوگرام: این روش به جای استفاده از تمام داده‌ها، از خلاصه‌های آماری داده‌ها استفاده می‌کند و به کاهش حافظه و پیچیدگی محاسباتی کمک می‌کند.

۲. بهینه‌سازی محاسبات: مانند PCA، استفاده از کتابخانه‌های محاسباتی که عملیات‌های ماتریسی را بهینه‌سازی می‌کنند، می‌تواند به کاهش زمان محاسبات کمک کند.

۳. کاهش ابعاد قبلی: قبل از اجرای LDA، استفاده از روش‌های کاهش ابعاد دیگر می‌تواند به کاهش بُعد داده‌ها و در نتیجه کاهش پیچیدگی محاسباتی کمک کند.

با توجه به این راه حل‌ها، می‌توان تا حدود زیادی مشکلات مرتبط با کاربرد PCA و LDA بر روی داده‌های با ابعاد بالا را کاهش داد و از این تکنیک‌های قدرتمند کاهش بُعد به نحو اثربخشی استفاده کرد.

ب)

اگر به جای ماتریس کوواریانس از معیارهای دیگری مانند اطلاعات متقابل برای ارزیابی ویژگی‌های داده‌ها استفاده کنیم، ممکن است به تغییرات قابل توجهی در نحوه تحلیل و تفسیر داده‌ها برسیم. در اینجا برخی از تغییرات مثبت و منفی احتمالی را مورد بررسی قرار می‌دهم:

جنبه‌های مثبت

۱. ارتباط مستقیم با کلاس‌ها: استفاده از اطلاعات متقابل می‌تواند به ما بینش‌های عمیق‌تری در مورد چگونگی تاثیر ویژگی‌ها بر لیبل‌ها بدهد. این امر می‌تواند به ویژه در زمینه‌هایی که دقت در تشخیص الگوهای خاص مهم است، ارزشمند باشد.

۲. کشف الگوهای پنهان: اطلاعات متقابل قابلیت شناسایی ارتباطات غیرخطی بین ویژگی‌ها و لیبل‌ها را دارد، که ممکن است توسط ماتریس کوواریانس نادیده گرفته شود.

جنبه‌های منفی

۱. پیچیدگی محاسباتی: تخمین اطلاعات متقابل معمولا نیاز به محاسبات بیشتری دارد، که می‌تواند به ویژه برای داده‌های بزرگ به یک چالش تبدیل شود.

۲. نیاز به داده‌های زیاد: برای اینکه تخمین اطلاعات متقابل دقیق باشد، نیاز به مجموعه داده‌های بزرگ است. در مجموعه داده‌های کوچک، خطر نتیجه‌گیری‌های نادرست وجود دارد.

۳. خطر بیش‌برازش: توجه بیش از حد به جزئیات موجود در داده‌های موجود می‌تواند منجر به مدل‌هایی شود که بر روی داده‌های جدید عملکرد خوبی ندارند.

در نهایت، انتخاب بین ماتریس کوواریانس و اطلاعات متقابل باید با در نظر گرفتن ویژگی‌های خاص مجموعه داده، اهداف تحلیلی، و منابع محاسباتی موجود صورت گیرد. در برخی موارد، ممکن است استفاده ترکیبی از هر دو رویکرد بهترین نتایج را به همراه داشته باشد.

پاسخ ۴.

سوال 4:

$$P(x) = \int P(x,z)\,dz = \int P(x|z)\,dz$$

$$\xrightarrow[\text{marginal}]{\text{توزیع نرمال}} P(x) = \int NP(z|0,1)\, Np(x|\mu + Wz, \sigma^2 I)\,dz = N(x|\mu, W^T W + \sigma^2 I)$$

یک راه دیگر برای نشان دادن نتیجه بالا این است که $x = \mu + Wz + \varepsilon$ باید نرمال باشد.

چون یک RV نرمال ← نرمال $\mu + Wz$ نرمال $P(z) = N P(z, 0, 1)$

ε نرمال $\rightarrow P(\varepsilon) = NP(\varepsilon|0, \sigma^2 I)$

در نهایت نتیجه می‌شود x نیز نرمال است.

$$E(x) = E[\mu + Wz + \varepsilon] = \mu + W \times 0 + 0 = \mu$$

$$\mathrm{Cov}[x] = E[(x-\mu)(x-\mu)^T] = E[(Wz + \varepsilon)(Wz + \varepsilon)^T] = WW^T + \sigma^2 I$$

ب)

$$P(z|x) = N(z|m, V) = N(z|W^T(WW^T + \sigma^2 I)^{-1}(x-\mu), I - W^T(WW^T + \sigma^2 I)^{-1} W)$$

پاسخ ۵.

سوال پنج:

الف)

$C_1 = \{x_1, x_2, x_3\}$ → $(0,2)$ $\{0,0\}$ $\{1.5,0\}$

$C_2 = \{x_3, x_5\}$ → $\{5,0\}$ $\{5,2\}$

برای C1 و C2 مرکز کلاستر را به دست می‌آوریم و سپس فاصله‌ی هر نقطه‌ای را حساب می‌کنیم تا فاصله از مرکز دسته‌ها به هر کس نزدیک‌تر بود به آن تعلق دارد من چند تا را می‌نویسم بقیه را در جدول حساب می‌کنم

$$X_1 = \left(\frac{0+0+5}{3}\right) = \frac{5}{3}, \quad X_2 = \left(\frac{2+0+0}{3}\right) = \frac{2}{3} \quad \Big\} C_1$$

$$X_1 = \left(\frac{1.5+5}{2}\right) = 3.25, \quad X_2 = \left(\frac{0+2}{2}\right) = 1 \quad \Big\} C_2$$

$$d(X_1, m_{C_1}) = \sqrt{\left(\frac{5}{3}-0\right)^2 + \left(\frac{2}{3}-2\right)^2} = \frac{\sqrt{41}}{3} = 2.13$$

$$d(X_1, m_{C_2}) = \sqrt{\left(\frac{6.5}{2}-0\right)^2 (1-2)^2} = \frac{\sqrt{46.25}}{2} = 3.4$$

نتیجه می‌شود که X_1 به دسته C_1 تعلق دارد

بقیه محاسبات را مانند بالا انجام می‌دهیم:

x_2 : $d(x_2, m_{c_1}) = 1.79$ $d(x_2, m_{c_2}) = 3.4$ مانند قبل به C_1 تعلق دارد

x_3 : $d(x_3, m_{c_1}) = 0.68$ $d(x_3, m_{c_2}) = 2.015$ مانند قبل به C_1 تعلق دارد

x_4 : $d(x_4, m_{c_1}) = 3.3$ $d(x_4, m_{c_2}) = 2.015$ مانند قبل به C_2 تعلق دارد

x_5 : $d(x_5, m_{c_1}) = 3.5$ $d(x_5, m_{c_2}) = 2.015$ مانند قبل به C_2 تعلق دارد

دوباره میانگین کلاستر ها آپدیت می‌شود اما واضح است همین باقی می‌ماند
پس کار همین جا پایان می‌یابد.

ب) برای این قسمت از Taxicab geometry یا فاصله‌ی منهتن استفاده می‌کنیم

$m_{c_1} = \left(\frac{5}{3}, \frac{2}{3}\right)$ $\quad m_{c_2} \ (3.25, 1)$ $\quad d(p,q) = \|p-q\|_1 = \sum_{i=1}^{n} |p_i - q_i|$

مثالی برای محاسبه برای x_2

$$d(x_2, m_{c_1}) = \left|\frac{5}{3} - 0\right| + \left|\frac{2}{3} - 0\right| = \frac{5}{3} + \frac{2}{3} = \frac{7}{3} = 2.33$$

نتیجه می‌شود x_2 به دسته‌ی اول تعلق دارد.

$$d(x_2, m_{c_1}) = \left|\frac{6.5}{2} - 0\right| + |1 - 0| = \frac{6.5}{2} + 1 = \frac{8.5}{2} = 4.25$$

بقیه مقادیر را به دست می‌آوریم و در جدول زیر می‌گذاریم

$x_1:$ $\quad d(x, m_{c_1}) = 3$ $\quad d(x, m_{c_2}) = 4.25$ مانند دسته‌بندی اولیه به C_1 تعلق دارد.

$x_3:$ $\quad d(x, m_{c_1}) = 0.84$ $\quad d(x, m_{c_2}) = 2.75$ مانند دسته‌بندی اولیه به C_1 تعلق دارد.

$x_4:$ $\quad d(x, m_{c_1}) = 4$ $\quad d(x, m_{c_2}) = 2.75$ مانند دسته‌بندی اولیه به C_2 تعلق دارد.

$x_5:$ $\quad d(x, m_{c_1}) = \frac{14}{3}$ $\quad d(x, m_{c_2}) = 2.75$ مانند دسته‌بندی اولیه به C_2 تعلق دارد.

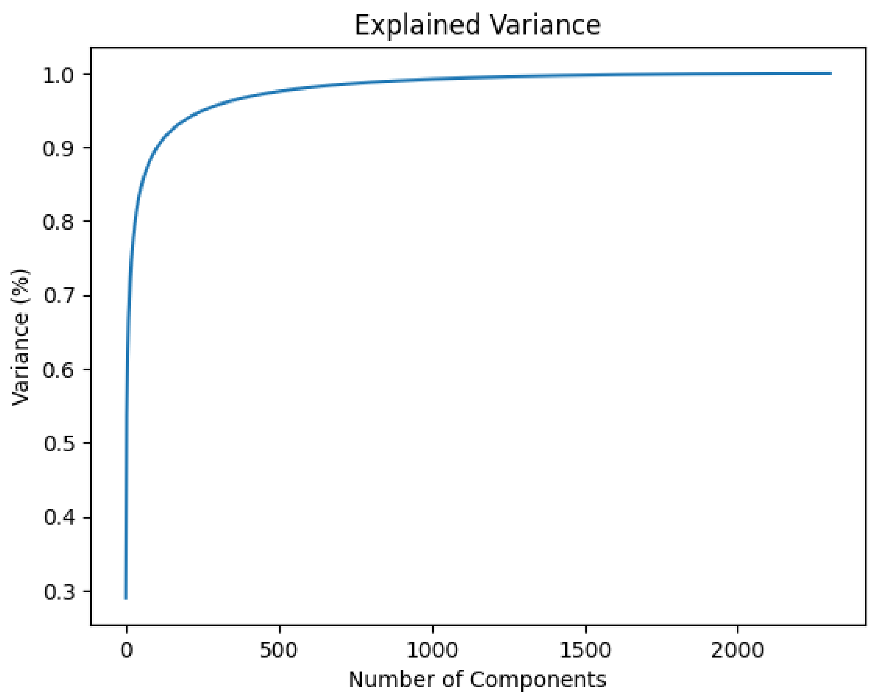
مثل سری قبل واضح است که دسته‌بندی تغییری نکرده است پس میانگین هم تغییر نمی‌کند برای هر کلاستر در نتیجه با تکرار الگوریتم باز به همین نتیجه خواهیم رسید و دسته‌بندی نهایی مثل قبل است.

$C_1 = \{x_1, x_2, x_3\}$

$C_2 = \{x_4, x_5\}$

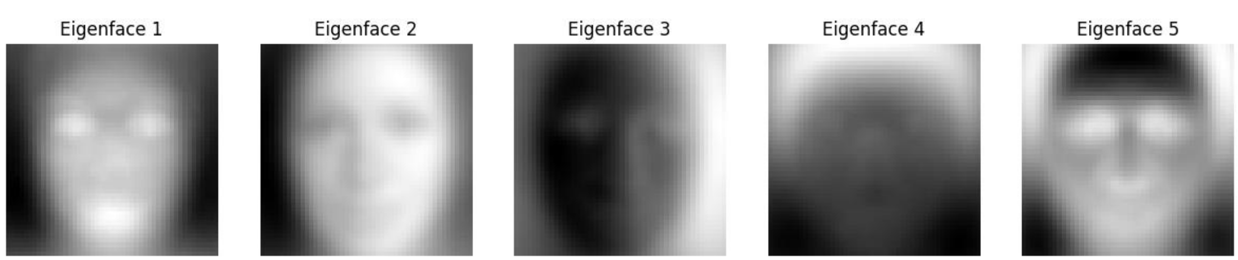
پاسخ ۶.

الف)Principal Component Analysis (PCA) یک تکنیک کاهش ابعاد است که برای فشرده‌سازی داده‌ها، به خصوص تصاویر، به کار می‌رود. در این فرآیند، مقادیر ویژه (Eigenvalues) از داده‌ها محاسبه می‌شوند که نشان‌دهنده میزان واریانسی هستند که هر کامپوننت (مولفه) جدید در داده‌ها توضیح می‌دهد.



در نمودار بالا، محور افقی تعداد کامپوننت‌ها و محور عمودی درصد واریانس توضیح داده شده توسط هر کامپوننت است. نقطه‌ای که منحنی شروع به تخت شدن می‌کند، می‌تواند به عنوان یک انتخاب خوب برای تعداد کامپوننت‌ها در نظر گرفته شود چرا که افزودن کامپوننت‌های بیشتر به طور قابل توجهی به میزان واریانس توضیح داده شده اضافه نمی‌کند.

ب)



این عکس‌ها نمایش دهنده بردارویژه (Eigenfaces) هستند که در پردازش تصویر و تشخیص چهره کاربرد دارند. ایگن‌فیس‌ها از تجزیه مقادیر منفرد یا تجزیه و تحلیل مؤلفه‌های اصلی (PCA) بر روی مجموعه‌ای از تصاویر چهره به دست می‌آیند و به عنوان یک مجموعه پایه برای توصیف چهره‌ها استفاده می‌شوند.

هر ایگن‌فیس نشان دهنده وزن‌هایی است که وقتی با میانگین چهره‌ها ترکیب می‌شود، می‌تواند تغییرات مختلف در نمای چهره انسان را تولید کند. به طور مثال:

- ایگن‌فیس ۱: ممکن است نمایانگر تغییرات عمده در روشنایی و کنتراست باشد.

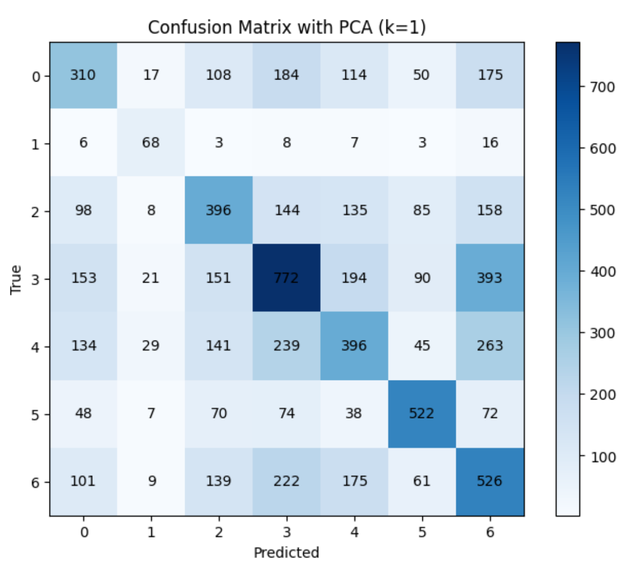
- ایگن‌فیس ۲: نشان دهنده ویژگی‌های مشخص‌تری مثل شکل کلی چهره، موقعیت چشم‌ها، بینی و دهان است.

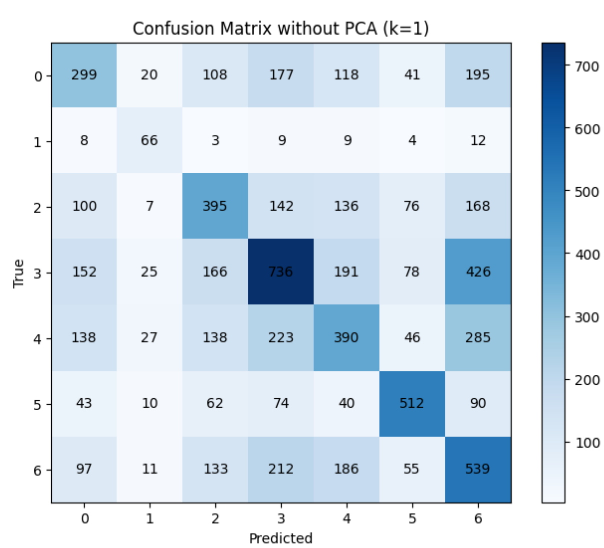
- ایگن‌فیس ۳: می‌تواند نشان دهنده تغییرات جهت‌دار نور و سایه‌های ناشی از آن باشد.

- ایگن‌فیس ۴ و ۵: ممکن است تغییرات کمتری را نشان دهند که بر ویژگی‌های دیگر چهره تأکید دارند، مانند بیان چهره یا ویژگی‌های خاص دیگر.

با ترکیب این ایگن‌فیس‌ها و اعمال وزن‌های مناسب، می‌توان تقریباً هر چهره‌ای را که در مجموعه داده وجود دارد بازسازی کرد. این روش در سیستم‌های تشخیص خودکار چهره برای شناسایی یا تأیید هویت استفاده می‌شود. همچنین می‌تواند برای فشرده‌سازی داده‌ها و کاهش ابعاد بدون از دست دادن اطلاعات مهم مورد استفاده قرار گیرد.

ج)





این دو تصویر ماتریس های درهم‌ریختگی (Confusion Matrix) را نشان می‌دهند که برای ارزیابی عملکرد مدل‌های تشخیص چهره با و بدون استفاده از تجزیه و تحلیل مؤلفه‌های اصلی (PCA) استفاده شده‌اند. در یک ماتریس درهم‌ریختگی، هر سطر نشان‌دهنده کلاس‌های واقعی و هر ستون نشان‌دهنده کلاس‌های پیش‌بینی شده توسط مدل است. اعداد درون سلول‌ها نشان‌دهنده تعداد نمونه‌هایی هستند که به آن کلاس تعلق دارند.

برای تحلیل و مقایسه این دو ماتریس:

۱. دقت کلی: می‌توان دقت کلی هر مدل را با محاسبه نسبت تعداد پیش‌بینی‌های صحیح (قرار گرفته در قطر اصلی) به تعداد کل پیش‌بینی‌ها ارزیابی کرد.

۲. خطاهای نوع اول و دوم: می‌توان خطاهای نوع اول (False Positives) و خطاهای نوع دوم (False Negatives) را برای هر کلاس محاسبه کرد تا ببینیم که مدل در کدام کلاس‌ها بیشتر دچار اشتباه می‌شود.

۳. قدرت تشخیصی: برای هر کلاس، می‌توان قدرت تشخیصی مدل را با بررسی نسبت تعداد پیش‌بینی‌های صحیح به کل نمونه‌های واقعی آن کلاس ارزیابی کرد.

- در هر دو ماتریس، قطر اصلی (جایی که کلاس‌های واقعی و پیش‌بینی شده مطابقت دارند) روشن‌ترین قسمت‌ها هستند، که نشان‌دهنده تعداد زیادی از پیش‌بینی‌های صحیح است.

- اختلافات عمده‌ای در تعداد پیش‌بینی‌های صحیح بین دو مدل وجود دارد، که می‌تواند نشان‌دهنده تأثیر PCA در کیفیت طبقه‌بندی باشد.

با مقایسه دقت کلی و معیارهای ارزیابی برای هر کلاس در مدل‌های با PCA و بدون PCA، می‌توانیم بگوییم:

۱. دقت کلی:

- مدل با PCA: دقت کلی تقریباً (۴۱.۶۶٪)

- مدل بدون PCA: دقت کلی تقریباً (۴۰.۹۲٪)

این نشان می‌دهد که استفاده از PCA کمی به بهبود دقت کلی مدل کمک کرده است.

۲. معیارهای ارزیابی برای هر کلاس (Precision, Recall و F1 Score):

برای هر کلاس، معیارهای ارزیابی مختلفی مانند دقت (Precision)، حساسیت (Recall) و FI Score را محاسبه کردیم. این معیارها برای هر کلاس به صورت زیر هستند:

- مدل با PCA:

- Precision: مقادیر بین (۳۶٪)-(۶۱٪)

- Recall: مقادیر بین (۳۲٪)-(۶۳٪)

- FI Score: مقادیر بین (۳۴٪)-(۶۲٪)

- مدل بدون PCA:

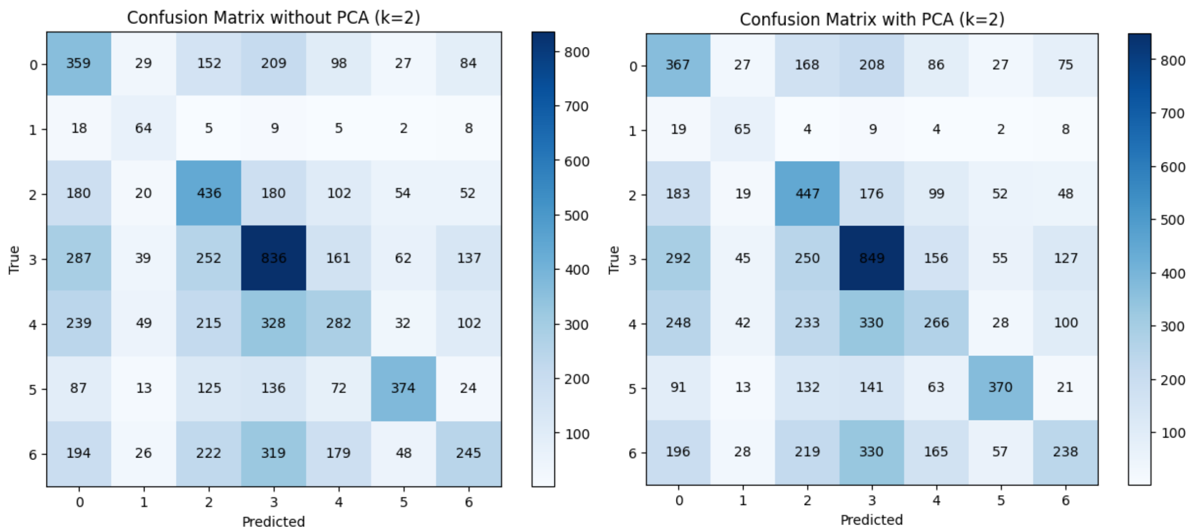
- Precision: مقادیر بین (۳۱٪)-(۶۳٪)

- Recall: مقادیر بین (۳۱٪)-(۶۲٪)

- FI Score: مقادیر بین (۳۳٪)-(۶۲٪)

با وجود تفاوت‌های جزئی، مدل با استفاده از PCA به طور کلی دقت بهتری در برخی کلاس‌ها نشان می‌دهد، اما تفاوت‌ها بسیار نزدیک هستند.

در مجموع، استفاده از PCA به نظر می‌رسد که به بهبود عملکرد کمک کرده است، اما تأثیر آن بسته به کلاس ممکن است متفاوت باشد. این تحلیل نشان می‌دهد که ممکن است برای برخی کلاس‌ها استفاده از PCA مفیدتر باشد، در حالی که برای دیگران تفاوت قابل توجهی ایجاد نکند.



با مقایسه معیارهای ارزیابی مدل‌ها با استفاده از PCA و بدون استفاده از PCA برای ($k = 2$):

۱. دقت کلی:

- مدل با PCA: دقت کلی تقریباً (٪۳۶.۲۵)

- مدل بدون PCA: دقت کلی تقریباً (٪۳۶.۱۷)

مشاهده می‌شود که دقت کلی دو مدل تقریباً مشابه است.

۲. معیارهای ارزیابی برای هر کلاس:

برای هر کلاس، معیارهای ارزیابی دقت (Precision)، حساسیت (Recall) و امتیاز (F1 Score) F1 به صورت زیر هستند:

- مدل با PCA:

- Precision: مقادیر بین (٪۲۶)-(٪۶۳)

- Recall: مقادیر بین (٪۱۹)-(٪۴۸)

- F1 Score: مقادیر بین (٪۲۶)-(٪۵۲)

- مدل بدون PCA:

- Precision: مقادیر بین (٪۲۶)-(٪۶۲)

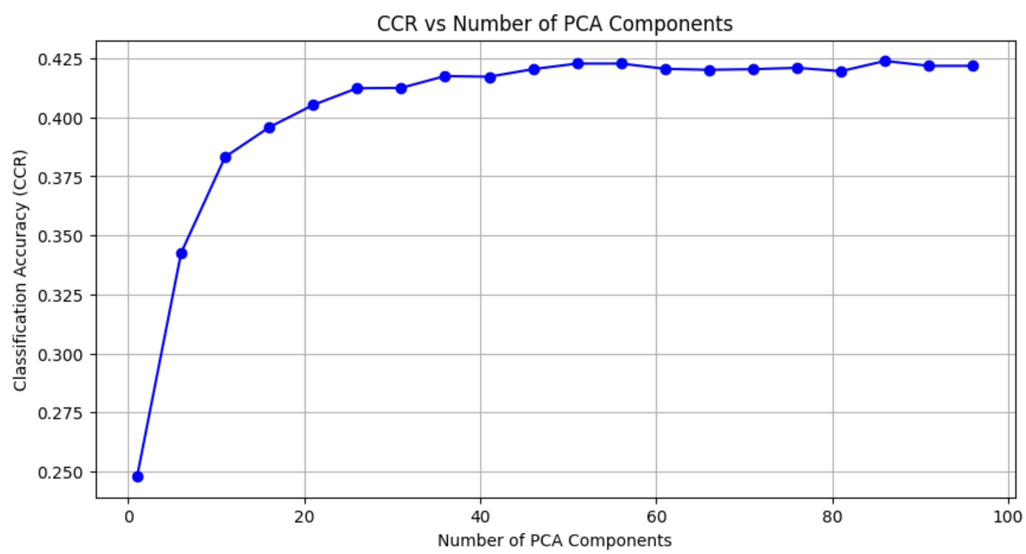
- Recall: مقادیر بین (٪۲۰)-(٪۴۷)

- FI Score: مقادیر بین (۲۶٪)-(۵۲٪)

هر دو مدل در کلاس‌های مختلف عملکرد نزدیکی به هم دارند و استفاده از PCA در این حالت تفاوت چشمگیری در معیارهای ارزیابی ایجاد نکرده است.

بنابراین، با افزایش تعداد ابعاد حفظ شده در PCA به ($k = 2$)، تفاوت عملکرد بین مدل‌ها با و بدون PCA کمتر شده و هر دو مدل عملکرد مشابهی دارند. این می‌تواند نشان دهد که برای این مجموعه داده و تعداد مؤلفه‌های انتخاب شده، PCA ارزش اضافی چشمگیری ایجاد نکرده است.

د)



این نمودار نشان‌دهنده رابطه بین تعداد مؤلفه‌های PCA و دقت طبقه‌بندی (CCR) برای یک مدل تشخیص چهره است. CCR مخفف Correct Classification Rate است که معیاری برای ارزیابی دقت یک مدل طبقه‌بندی است. نکات مهمی که از نمودار استخراج می‌شوند عبارتند از:

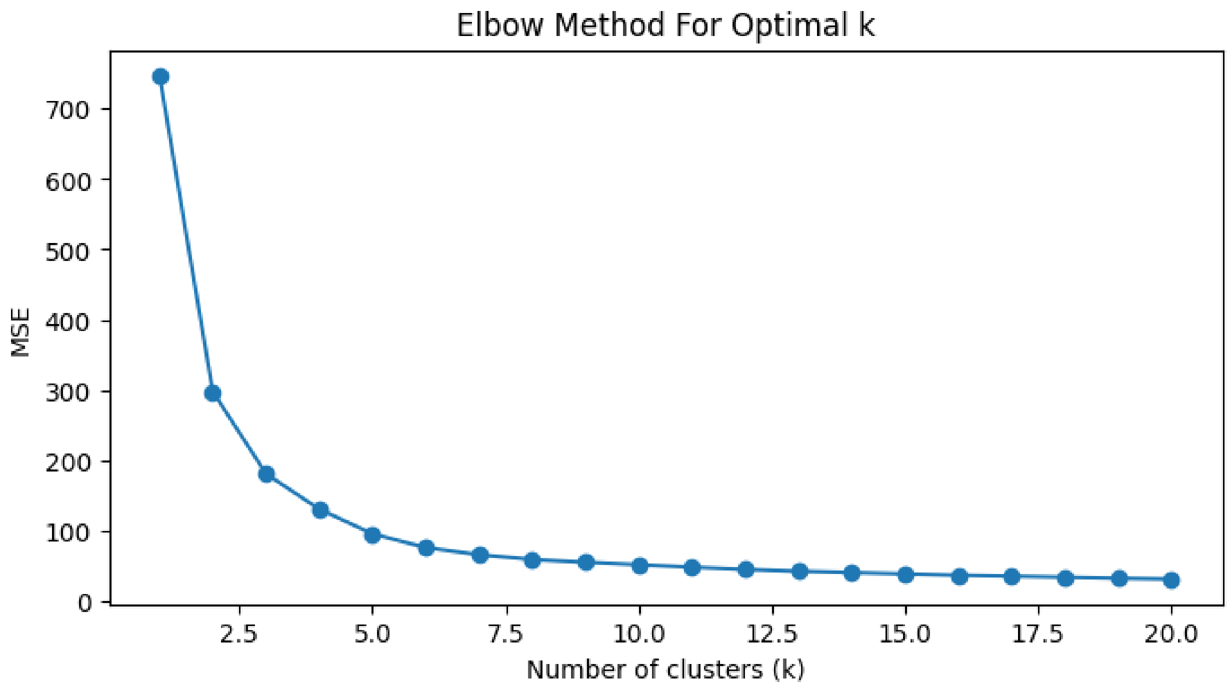
- افزایش اولیه دقت: با افزایش تعداد مؤلفه‌های PCA از ۰ به ۲۰، یک افزایش چشمگیر در دقت طبقه‌بندی مشاهده می‌شود. این نشان می‌دهد که مؤلفه‌های اصلی اولیه حاوی بیشترین اطلاعات مفید برای تشخیص چهره هستند.

- تخت شدن نمودار: پس از حدود ۲۰ مؤلفه، افزایش دقت طبقه‌بندی کندتر می‌شود و نمودار شروع به تخت شدن می‌کند. این نشان می‌دهد که اضافه کردن مؤلفه‌های بیشتر تأثیر کمتری بر بهبود دقت دارد.

- پلاتوی دقت: در حدود ۴۰ مؤلفه، دقت به یک پلاتو می‌رسد و تقریباً تغییری نمی‌کند، حتی با افزودن مؤلفه‌های بیشتر تا ۱۰۰. این می‌تواند به این معنی باشد که افزودن مؤلفه‌های بیشتر پس از این نقطه اطلاعات اضافی قابل توجهی به مدل اضافه نمی‌کند و ممکن است فقط باعث پیچیدگی بیشتر و افزایش محاسبات شود بدون اینکه دقت قابل توجهی اضافه شود.

این تحلیل می‌تواند به ما کمک کند تا تعداد مؤلفه‌های PCA را برای بهینه‌سازی دقت و کارایی مدل انتخاب کنیم. در این مثال، ممکن است بین ۲۰ تا ۴۰ مؤلفه یک نقطه بهینه وجود داشته باشد که با حفظ دقت قابل قبول، هزینه‌های محاسباتی را کاهش می‌دهد.

پاسخ ۷.



در نمودار، محور افقی تعداد خوشه‌ها (`k`) را نشان می‌دهد و محور عمودی خطای مربعات میانگین (MSE) را نشان می‌دهد. هدف این است که تعداد `k` را به‌گونه‌ای انتخاب کنیم که با افزایش تعداد خوشه‌ها، کاهش قابل‌توجهی در MSE را شاهد نباشیم. به عبارت دیگر، به دنبال نقطه‌ای هستیم که افزودن خوشه‌های بیشتر باعث بهبود قابل‌توجهی در کاهش خطا نشود.

بر اساس نمودار، می‌بینیم که با افزایش `k` از ۱ به ۲، یک کاهش شدید در MSE وجود دارد. با ادامه این روند، افزایش `k` از ۲ به ۳ و بیشتر باعث کاهش‌های کمتر و کمتری در MSE می‌شود. به نظر می‌رسد که نقطه آرنج یا تعداد بهینه‌ی `k` در این نمودار بین ۲ تا ۸ باشد، زیرا پس از آن، کاهش MSE کمتر می‌شود و نمودار به مرور صاف‌تر می‌شود.

تعیین دقیق نقطه آرنج ممکن است به تجزیه و تحلیل بیشتر یا انتخاب‌های موضوعی نیاز داشته باشد.

Compressed image

Original image

```
Original size:  (650, 488)
Compressed size:  488.64385618977474
```

نتایج به صورت بالا است.