

بسم الله الرحمن الرحيم



دانشکده مهندسی برق و کامپیوتر

یادگیری ماشین - تمرین پنجم

سید مهدی رضوی

استاد : آقای دکتر توسلی پور - آقای دکتر ابوالقاسمی

۱۴۰۳ تیر ماه



فهرست مطالب

۳	۱ تمرین اول
۵	۲ تمرین دوم
۸	۳ تمرین سوم
۹	۴ تمرین چهارم
۱۰	۵ تمرین پنجم
۱۴	۶ تمرین ششم
۱۷	۷ تمرین هفتم

فهرست تصاویر

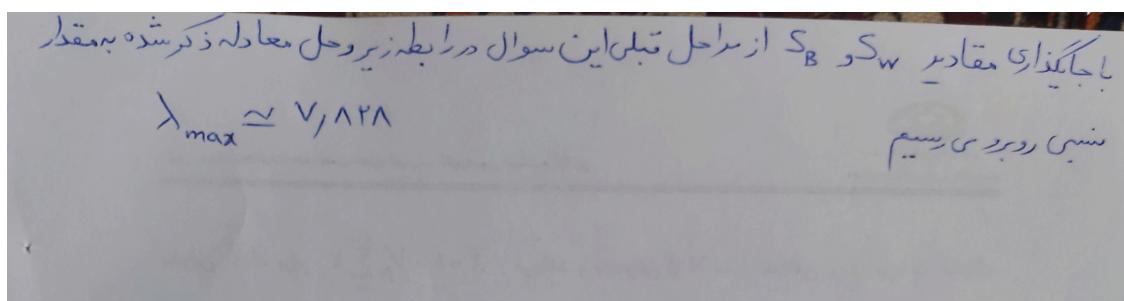
۳	پاسخ سوال اول	۱
۴	پاسخ سوال اول	۲
۸	پاسخ سوال سوم	۳
۱۰	مشاهده یک نمونه از مجموعه داده مدنظر سوال	۴
۱۱	Heatmap of covariance matrix	۵
۱۱	Heatmap of covariance matrix 2	۶
۱۱	نمایش مقادیر ویژه به صورت نزولی	۷
۱۲	بررسی احتمال برای تجمع تعداد مولفه‌ها و بررسی احتمال اطمینان برای الگوریتم کاهش ابعاد PCA	۸
۱۲	اعمال الگوریتم PCA بر روی مجموعه داده	۹
۱۲	مشاهده یک نمونه از مجموعه داده و بازسازی شده آن	۱۰
۱۳	مشاهده یک نمونه از مجموعه داده و بازسازی شده آن	۱۱
۱۴	اعمال الگوریتم کاهش ابعاد بر روی مجموعه داده مدنظر سوال	۱۲
۱۵	مدل مخلوط گوسی بر روی مجموعه داده به همراه نقاط میانگین آنها	۱۳
۱۵	بازسازی تصویر از روی نقاط کاهش بعد داده شده	۱۴
۱۶	نمونه‌های با کمترین اختلاف در احتمال تعلق به کلاس‌ها	۱۵
۱۶	کمترین و بیشترین اختلاف بین نمونه‌ها	۱۶
۱۸	انواع نمودارهای ذکر شده برای خوشبندی نقاط داده مشتریان	۱۷
۱۸	شش خوش برای داده مشتریان بر اساس دو ویژگی سن و حقوق دریافتی	۱۸
۱۹	نمودارهای مختلف برای شش خوش	۱۹
۱۹	نمودارهای مختلف برای شش خوش	۲۰

۱ تمرین اول

$$\begin{aligned}
 S_w &= S_{w_1} + S_{w_2} & \mu_1 &= \frac{1}{\Delta} \begin{bmatrix} 1\Delta \\ 1\Delta \end{bmatrix} = \begin{bmatrix} 3 \\ 3/2 \end{bmatrix} \\
 S_w &= \sum_{i=1}^N (x_i - \mu_i) (x_i - \mu_i)^T & \mu_2 &= \frac{1}{\Delta} \begin{bmatrix} \varepsilon 2 \\ 3\Delta \end{bmatrix} = \begin{bmatrix} \varepsilon/2 \\ 1/2 \end{bmatrix} \\
 S_{w_1} &= \begin{bmatrix} 1 \\ -1/2 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ 1/2 & 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ -1/2 & 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1/2 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} \\
 &\quad + \begin{bmatrix} 0 \\ 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0/2 \\ 0/2 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1+1+1+0+1 & -1/2-1/2+1/2+0+0 \\ -1/2-1/2+0/2+0+1/2 & 1/2+1/2+1/2+1/2+1/2 \end{bmatrix} = \begin{bmatrix} \varepsilon & -1/2 \\ -1/2 & 13/2 \end{bmatrix} \\
 S_{w_2} &= \begin{bmatrix} 0/2 \\ 1/2 \end{bmatrix} \begin{bmatrix} 0/2 & 1/2 \\ 1/2 & 0/2 \end{bmatrix} + \begin{bmatrix} -1/2 \\ 0/2 \end{bmatrix} \begin{bmatrix} -1/2 & 0/2 \\ 0/2 & -1/2 \end{bmatrix} + \begin{bmatrix} 0/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} 0/2 & -1/2 \\ -1/2 & 0/2 \end{bmatrix} + \\
 &\quad \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 0/2 \end{bmatrix} \begin{bmatrix} 1/2 & 0/2 \\ 0/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 9/2 & -1/2 \\ -1/2 & 13/2 \end{bmatrix} \\
 S_w &= \begin{bmatrix} 13/2 & -1/2 \\ -1/2 & 25/2 \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
 m_1 &= \frac{1}{\Delta} \begin{bmatrix} \varepsilon + 2 + 2 + 3 + \varepsilon \\ 1 + \varepsilon + 3 + 7 + \varepsilon \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 1\Delta \\ 1\Delta \end{bmatrix} = \begin{bmatrix} 3 \\ 3/2 \end{bmatrix} & \text{سؤال ۱} \\
 m_2 &= \frac{1}{\Delta} \begin{bmatrix} 9 + 9 + 9 + 1 + 10 \\ 10 + 1 + 3 + 5 + 1 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \varepsilon 2 \\ 3\Delta \end{bmatrix} = \begin{bmatrix} \varepsilon/2 \\ 1/2 \end{bmatrix} \\
 S_B &= (m_1 - m_2)(m_1 - m_2)^T \\
 m_1 - m_2 &= \begin{bmatrix} 3 - 1/2 \\ 3/2 - 1/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix} & (m_1 - m_2)^T &= \begin{bmatrix} -1/2 & -1/2 \end{bmatrix} \\
 S_B &= \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} -1/2 & -1/2 \end{bmatrix} = \begin{bmatrix} 29/16 & 21/16 \\ 21/16 & 16 \end{bmatrix}
 \end{aligned}$$

شکل ۱: پاسخ سوال اول



شكل ۲: پاسخ سوال اول



۲ تمرین دوم

۱. model selection

انتخاب مدل شامل انتخاب بهترین مدل یا الگوریتم از بین مجموعه‌ای از مدل‌های کاندید است. این فرآیند معمولاً شامل مقایسه عملکرد مدل‌های مختلف بر روی یک مجموعه داده است. مقایسه دقت، Recall و F1 Score خواهد بود.

انتخاب مدل در طول فاز آموزش یک پروژه یادگیری ماشین انجام می‌شود.

۲. model assessment

ارزیابی مدل از سوی دیگر شامل ارزیابی عملکرد یک مدل انتخاب شده بر روی داده‌های ناشناخته برای ارزیابی توانایی عمومی آن است. این کار معمولاً با استفاده از یک مجموعه داده آزمایش جداگانه که در طول آموزش مدل استفاده نشده است، انجام می‌شود. ارزیابی مدل به کمک تخمین زدن کمک می‌کند که چقدر مدل احتمالاً در حالت‌های واقعی عمل خواهد کرد.

به طور خلاصه، انتخاب مدل برای انتخاب بهترین مدل از بین گزینه‌های کاندید و ارزیابی مدل برای ارزیابی عملکرد مدل انتخاب شده بر روی داده‌های ناشناخته استفاده می‌شود. هر دو فرآیند برای توسعه موثر مدل‌های یادگیری ماشین حائز اهمیت هستند.

دلایل استفاده از انتخاب مدل شامل :

۱. یافتن بهترین مدل برای یک مجموعه داده خاص

۲. پیشگیری از بیشبرازش OverFitting

۳. بهبود عملکرد مدل

۴. صرفه‌جویی در طول مدت زمان آموزش و منابع سخت‌افزاری مانند GPU از سایتها که به صورت محدود منابع در اختیار ما قرار می‌دهند.



الگوریتم‌های انتخاب مدل احتمالاتی و الگوریتم‌های انتخاب مدل با استفاده از نمونه‌برداری دو روش مختلف برای انتخاب بهترین مدل از بین مجموعه‌های از مدل‌های کاندید هستند. در ادامه، یک مقایسه بین این دو روش ارائه می‌شود:

۱. الگوریتم‌های انتخاب مدل احتمالاتی: - الگوریتم‌های انتخاب مدل احتمالاتی از روش‌های آماری برای برآورد احتمال مدل با توجه به داده‌ها استفاده می‌کنند. - این الگوریتم‌ها معمولاً شامل محاسبه احتمال داده تحت هر مدل و استفاده از استنباط بیزی برای مقایسه مدل‌ها هستند. - نمونه‌هایی از الگوریتم‌های انتخاب مدل احتمالاتی شامل BIC (Bayesian Information Criterion) و DIC (Deviance Information Criterion) AIC (Akaike Information Criterion) هستند. - این الگوریتم‌ها یک روش اصولی برای مقایسه مدل‌ها بر اساس پیچیدگی و خوبی مناسب با داده‌ها فراهم می‌کنند.

۲. الگوریتم‌های انتخاب مدل با استفاده از نمونه‌برداری: - الگوریتم‌های انتخاب مدل با استفاده از نمونه‌برداری شامل نمونه‌برداری مقابله و بوت استرپینگ است که به صورت تکراری نمونه‌برداری از زیرمجموعه‌های داده برای ارزیابی عملکرد مدل‌های مختلف استفاده می‌شود. - تکنیک‌های رایج نمونه‌برداری شامل اعتبارسنجی مقابله و بوت استرپینگ هستند. - اعتبارسنجی مقابله شامل تقسیم داده به دسته‌های آموزش و آزمون به صورت تکراری برای برآورد عملکرد مدل بر روی داده‌های نامشخص است. - بوت استرپینگ شامل ایجاد چندین نمونه بوت استرپ از داده اصلی و استفاده از آن‌ها برای ارزیابی پایداری عملکرد مدل است. مقایسه: - الگوریتم‌های انتخاب مدل احتمالاتی بر روی اصول آماری و برآورد احتمال تخصص دارند، در حالی که الگوریتم‌های انتخاب مدل با استفاده از نمونه‌برداری از روش‌های تجربی مانند اعتبارسنجی مقابله و بوت استرپینگ استفاده می‌کنند. - الگوریتم‌های احتمالاتی براساس دیدگاه‌های نظری پیرامون پیچیدگی مدل و خوبی مناسب با داده، در حالی که الگوریتم‌های نمونه‌برداری یک راه عملی برای برآورد عملکرد کلی مدل با شبیه سازی فرآیند آموزش و آزمون بر روی زیرمجموعه‌های مختلف داده فراهم می‌کنند. - الگوریتم‌های احتمالاتی ممکن است به لحاظ محاسباتی چالش برانگیز تر باشند زیرا شامل محاسبات احتمال و استفاده از استنباط بیزی هستند، در حالی که الگوریتم‌های نمونه‌برداری آسان‌تر در پیاده سازی و تفسیر هستند. - هر دو نوع الگوریتم قوا و ضعف خود را دارند و انتخاب بین آن‌ها ممکن است به نیازهای خاص وظایف مدلسازی و منابع محاسباتی در دسترس بستگی داشت.

بطور خلاصه، الگوریتم‌های انتخاب مدل احتمالاتی براساس اصول آماری و برآورد احتمال، در حالی که الگوریتم‌های نمونه‌برداری با استفاده از نمونه‌برداری روش‌های تجربی استفاده کرده و چارچوب عملکرد کلان مدل را به صورت تجربی تخمین زده است. انتخاب بین این دو روش به نیاز خاص وظیفة مدلسازی و منابع محاسباتی در دسترس بستگی دارد.



وقتی داده‌ها کم است، انتخاب مدل به دلیل محدود بودن اطلاعات موجود برای تخمین دقیق عملکرد مدل‌های مختلف چالش‌برانگیز می‌شود. در زیر چالش‌ها و راهکارهایی برای انتخاب و ارزیابی مدل در شرایط داده کم آورده شده است:

چالش‌ها در این هنگام عبارت است از :

۱. OverFitting

۲. تخمین انحرافی از عملکرد یا پارامترهای مدل

۳. عدم قطعیت در انتخاب مدل

۴. ممکن است سخت باشد که نتایج یک مجموعه داده کوچک را به جمعیت بزرگتر تعمیم داد.

راهکارهای انتخاب و ارزیابی مدل :

۱. مدل‌های ساده‌تر

k fold validation . ۲

۳. روش‌های مختلف نمونه‌برداری مثل Bootstrapping

۴. اضافه‌کردن عبارت منظم‌ساز L1 Lasso or L2 Lasso

به طور خلاصه، زمانی که با دسترسی کم به داده سروکار داریم، حواس خود را به پیچیدگی مدل، استفاده از تکنیک‌های اعتبارسنجی قوی و در نظر گرفتن دانش حوزه یا روش‌های منظم‌سازی جلب می‌کنیم تا چالش‌های انتخاب و ارزیابی مدل را به خوبی حل کنیم.



٣ تمرین سوم

سؤال ٣

$$P(x) = \alpha \lambda_1 e^{-\lambda_1 x} + (1-\alpha) \lambda_2 e^{-\lambda_2 x}$$

$$P(x_i) = \alpha \lambda_1 e^{-\lambda_1 x_i} + (1-\alpha) \lambda_2 e^{-\lambda_2 x_i}$$

DataSet $\{x_1, x_2, \dots, x_n\}$

$$L(\alpha, \lambda_1, \lambda_2; x_1, x_2, \dots, x_n) = \prod_{i=1}^n \left[\frac{\alpha \lambda_1 e^{-\lambda_1 x_i}}{(1-\alpha) \lambda_2 e^{-\lambda_2 x_i}} \right]$$

$$\log L(\alpha, \lambda_1, \lambda_2; x_1, x_2, \dots, x_n) = \sum_{i=1}^n \log \left[\frac{\alpha \lambda_1 e^{-\lambda_1 x_i}}{(1-\alpha) \lambda_2 e^{-\lambda_2 x_i}} \right]$$

Complete Log-likelihood

E-Step :

$$\begin{cases} w_{i1} = \frac{\alpha \lambda_1 e^{-\lambda_1 x_i}}{\alpha \lambda_1 e^{-\lambda_1 x_i} + (1-\alpha) \lambda_2 e^{-\lambda_2 x_i}} \\ w_{i2} = \frac{(1-\alpha) \lambda_2 e^{-\lambda_2 x_i}}{\alpha \lambda_1 e^{-\lambda_1 x_i} + (1-\alpha) \lambda_2 e^{-\lambda_2 x_i}} \end{cases}$$

$$w_{i1} + w_{i2} = 1$$

M-Step :

$$\alpha = \frac{1}{n} \sum_{i=1}^n w_{i1}$$

$$\lambda_1 = \frac{\sum_{i=1}^n w_{i1}}{\sum_{i=1}^n w_{i1} x_i}$$

Maximize the
Expected Parameters

~~$$\lambda_2 = \frac{\sum_{i=1}^n w_{i2}}{\sum_{i=1}^n w_{i2} x_i}$$~~

شكل ٣: پاسخ سوال سوم



۴ تمرین چهارم

ابتدا باید در نظر داشته باشیم که مدل GMM متشکل از مثلا k تا مدل توزیع گوسی خواهد بود.
پس ما باید به همین تعداد

 π_i μ_i Σ_i

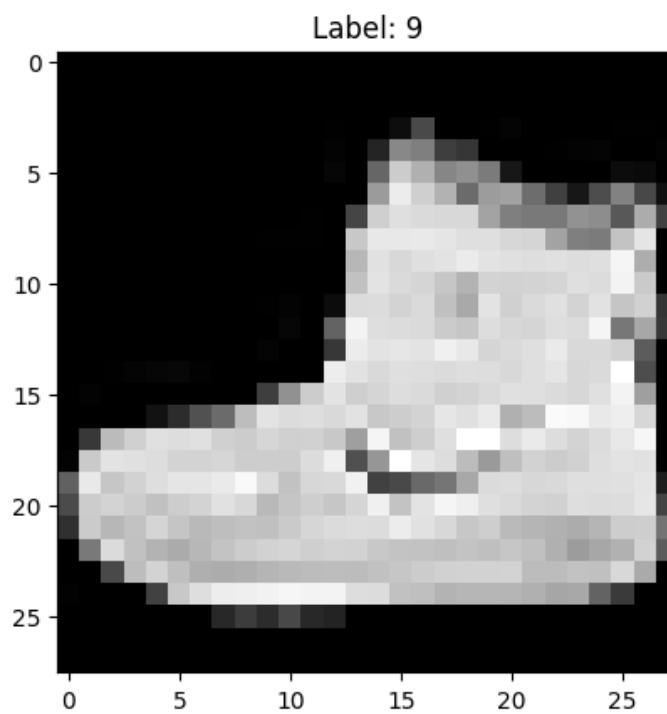
در نتیجه باید به اندازه $3k$ پارامتر را شبکه عصبی ما آموزش ببینند.

از تابع Softmax برای فعال‌سازی استفاده خواهیم کرد که ضرایب را آموزش ببینند.

برای محاسبه میانگین‌ها و ماتریس واریانس‌ها می‌توانیم از تابع فعال‌سازی خطی نیز استفاده کنیم.
تابع خطأ برای آموزش شبکه به صورت زیر خواهد بود.

$$L(\theta) = -\sum_{i=1}^n (\sum_{k=1}^K \pi_k N(x_i | \mu_k, \Sigma_k)) \quad (1)$$

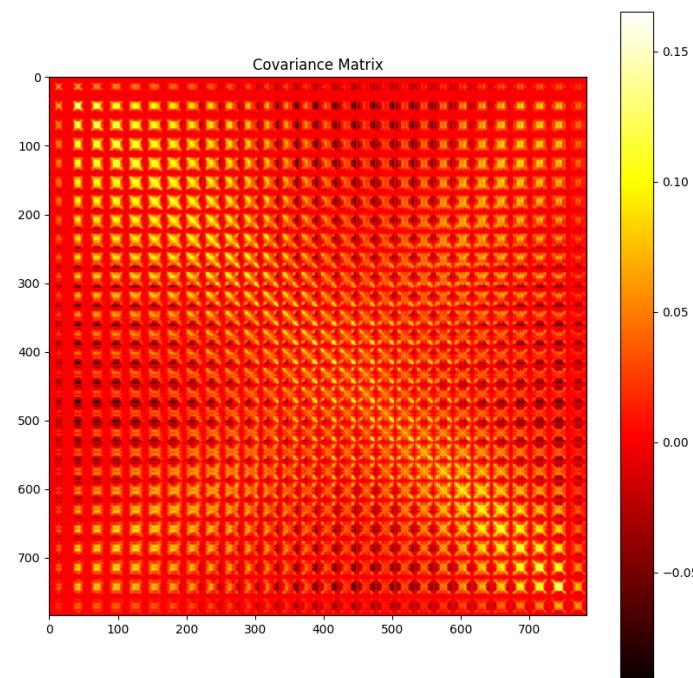
۵ تمرین پنجم



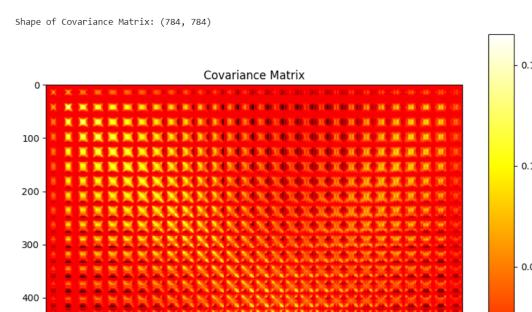
شکل ۴: مشاهده یک نمونه از مجموعه داده مدنظر سوال

ابعاد ماتریس کوواریانس یک مربع به اندازه بردار ویژگی در اندازه بردار ویژگی خواهد بود.

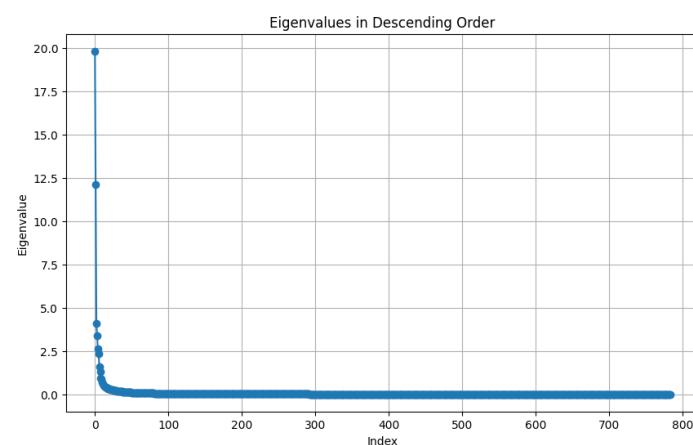
Shape : (784 * 784)



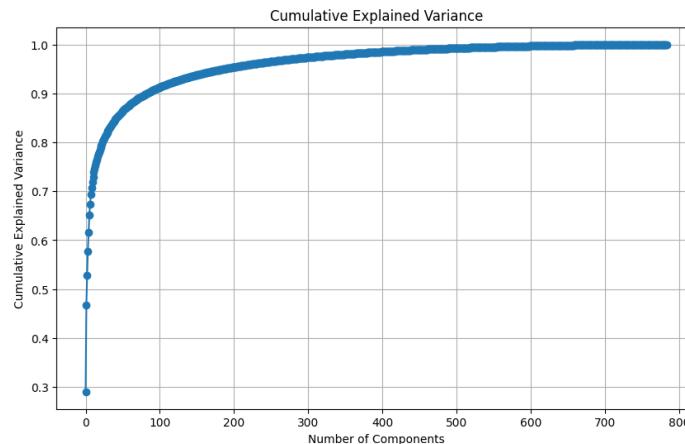
شکل ۵: Heatmap of covariance matrix



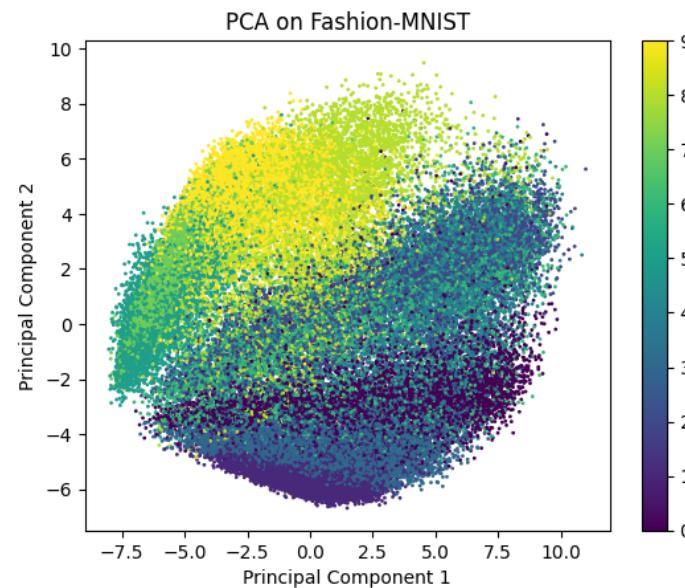
شکل ۶: Heatmap of covariance matrix 2



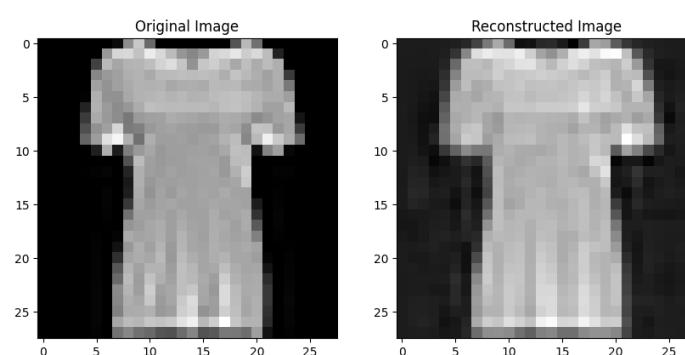
شکل ۷: نمایش مقادیر ویژه به صورت نزولی



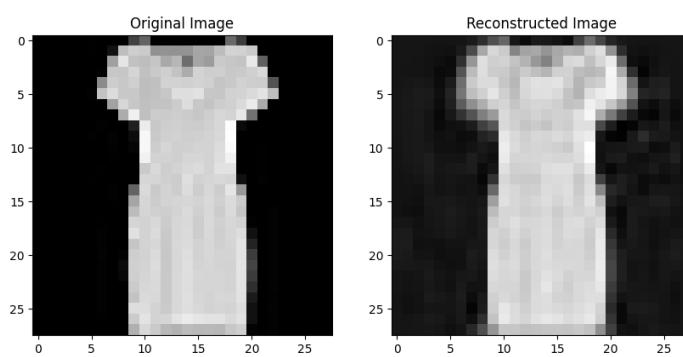
شکل ۸: بررسی احتمال برای تجمع تعداد مولفه‌ها و بررسی احتمال اطمینان برای الگوریتم کاهش ابعاد PCA



شکل ۹: اعمال الگوریتم PCA بر روی مجموعه داده

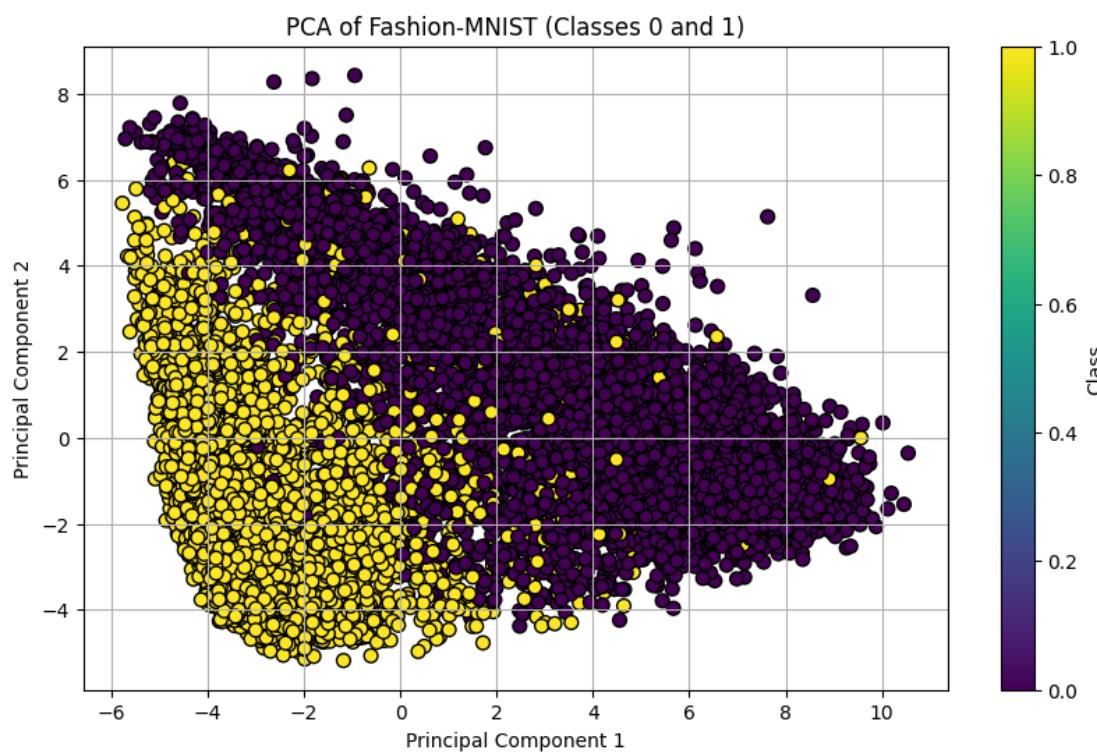


شکل ۱۰: مشاهده یک نمونه از مجموعه داده و بازسازی شده آن



شکل ۱۱: مشاهده یک نمونه از مجموعه داده و بازسازی شده آن

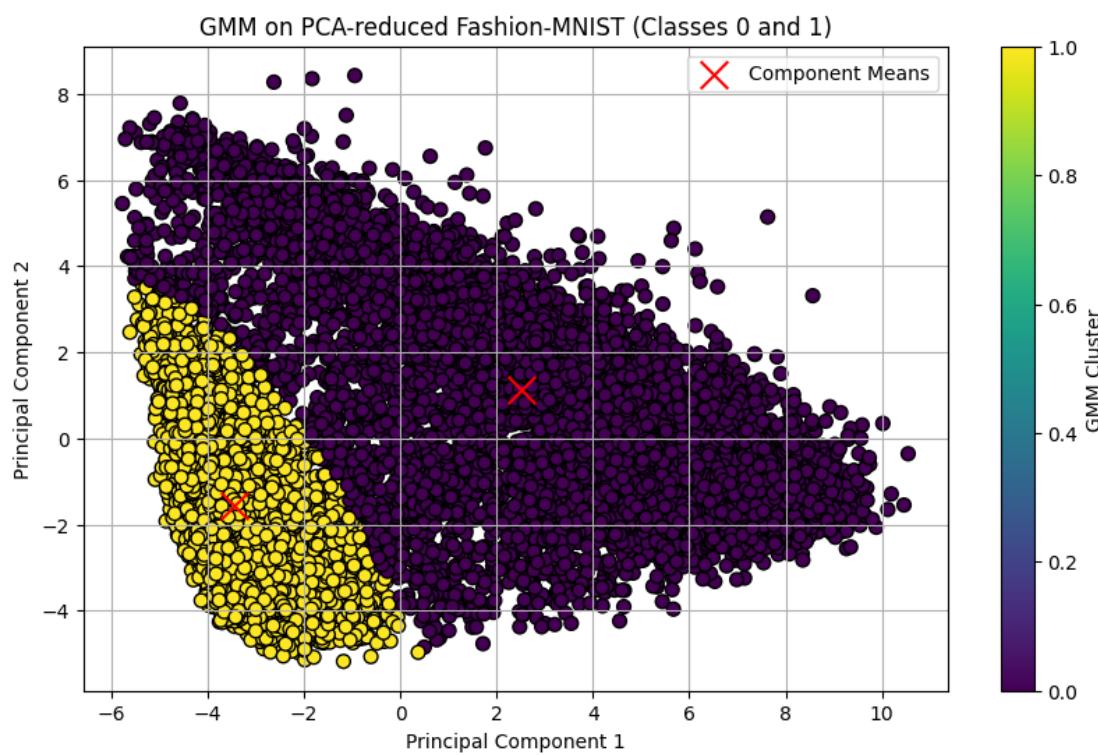
۶ تمرین ششم



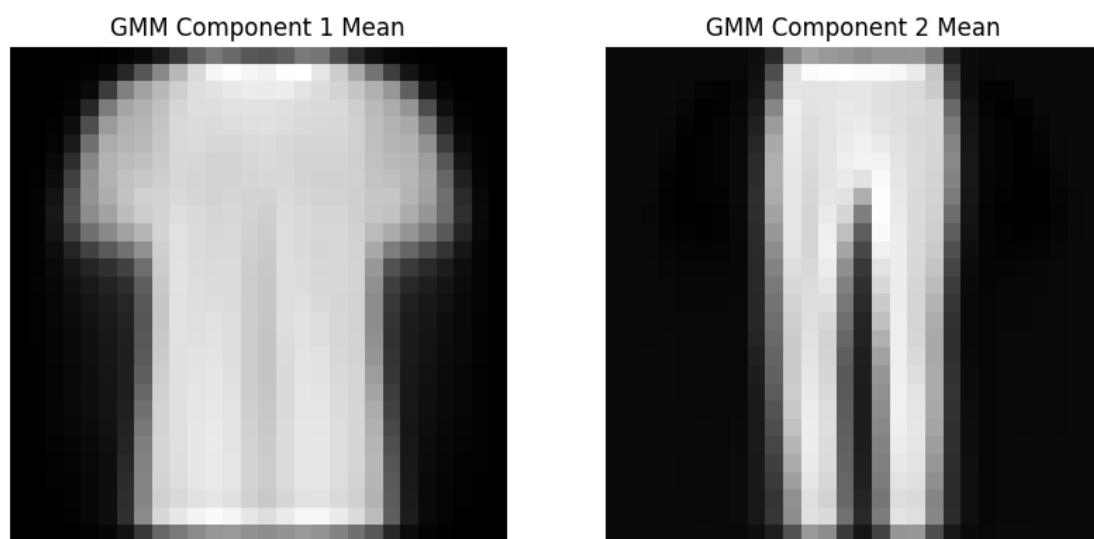
شکل ۱۲: اعمال الگوریتم کاهش ابعاد بر روی مجموعه داده مدنظر سوال

فاصله اقلیدسی بین میانگین نقاط دو کلاس 6.5491012156196255 می‌باشد.

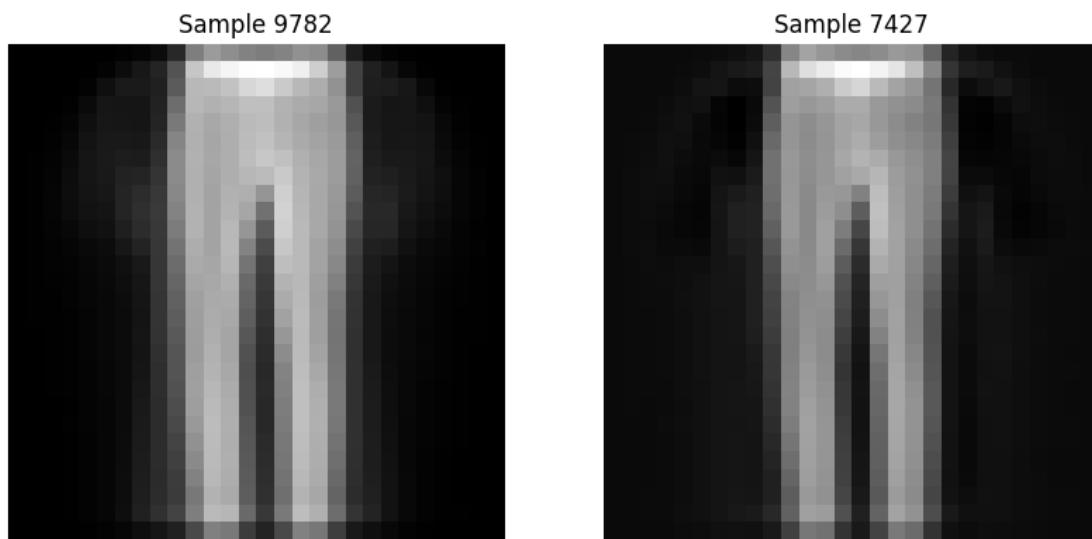
همانطور که از تصویر بالا نیز مشخص است، الگوریتم نمی‌تواند تمام فضای فیچر را از روی داده کاهش بعد یافته بازسازی کند و تنها نقاط اصلی تصویر بازسازی شده است. این بدان معناست که ویژگی‌های اصلی تصویر بازسازی شده است.



شکل ۱۳: مدل مخلوط گوسی بر روی مجموعه داده به همراه نقاط میانگین آن‌ها



شکل ۱۴: بازسازی تصویر از روی نقاط کاهش بعد داده شده



شکل ۱۵: نمونه‌های با کمترین اختلاف در احتمال تعلق به کلاس‌ها

```
max_distance, max_pair = max(distances, key=lambda x: x[0])
min_distance, min_pair = min(distances, key=lambda x: x[0])

print(f'Largest distance: {max_distance} between classes {max_pair}')
print(f'Smallest distance: {min_distance} between classes {min_pair}')
```

→ Largest distance: 10.70480391166218 between classes (1, 9)
Smallest distance: 4.313038351319221 between classes (5, 7)

شکل ۱۶: کمترین و بیشترین اختلاف بین نمونه‌ها



۷ تمرین هفتم

۱ K-Means Distortion .

یک الگوریتم محبوب خوشبندی است که هدف آن تقسیم یک مجموعه داده به K خوش است، که K یک پارامتر تعریف شده توسط کاربر است. نمودار تنوع K-Means نشان می‌دهد مجموع مریع فواصل بین هر نقطه داده و مرکز خوش اختصاص یافته به آن، به عنوان یکتابع از K است. ایده این است که مقدار K را انتخاب کنید که تنوع را به حداقل برساند، که یک اندازه‌گیری از چقرومگی خوش‌ها است.

۲. امتیاز سیلوئت: امتیاز سیلوئت یک اندازه‌گیری است که نشان می‌دهد هر نقطه داده چقدر به خوش اختصاص یافته به آن می‌خورد، نسبت به سایر خوش‌ها. این امتیاز از ۱ - تا ۱ متغیر است، جایی که یک امتیاز ۱ نشان دهنده این است که نقطه داده به خوبی به خوش اختصاص یافته است، و یک امتیاز ۱ - نشان دهنده این است که بهتر بود در یک خوش دیگر قرار می‌گرفت. نمودار امتیاز سیلوئت میانگین امتیاز برای هر مقدار K را نشان می‌دهد و می‌تواند برای انتخاب تعداد بهینه خوش‌ها استفاده شود.

۳. شاخص Davies-Bouldin .

شاخص Davies-Bouldin یک اندازه‌گیری دیگر از کیفیت خوشبندی است، براساس میانگین فاصله بین هر مرکز خوش و نزدیکترین همسایه خود است. این شاخص از ۰ تا بینهایت متغیر است، جایی که امتیاز پایین‌تر بهترین خوشبندی را نشان می‌دهد. نمودار شاخص Davies-Bouldin نمره برای هر مقدار K را نشان می‌دهد و می‌تواند برای انتخاب تعداد بهینه خوش‌ها استفاده شود.

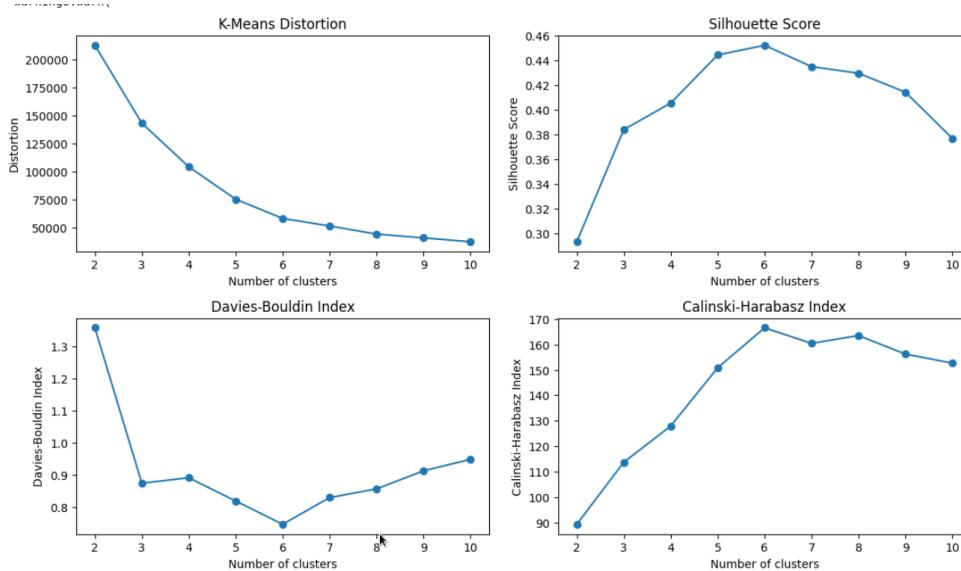
۴. شاخص Calinski-Harabasz .

شاخص Calinski-Harabasz یک اندازه‌گیری است که نشان می‌دهد چقدر خوش‌ها از هم جدا هستند، براساس نسبت واریانس بین خوش به واریانس داخل خوش. این شاخص از ۰ تا بینهایت متغیر است، جایی که امتیاز بالاتر بهترین خوشبندی را نشان می‌دهد. نمودار شاخص Calinski-Harabasz نمره برای هر مقدار K را نشان می‌دهد و می‌تواند برای انتخاب تعداد بهینه خوش‌ها استفاده شود.

۵. شاخص Dunn .

شاخص Dunn یک اندازه‌گیری دیگر از جدایی خوش‌ها است، براساس نسبت حداقل فاصله بین هر دو نقطه داده در دو خوش مختلف به قطر بزرگترین خوش. این شاخص از ۰ تا بینهایت متغیر است، جایی که امتیاز بالاتر بهترین خوشبندی را نشان می‌دهد. نمودار شاخص Dunn نمره برای هر مقدار K را نشان می‌دهد و می‌تواند برای انتخاب تعداد بهینه خوش‌ها استفاده شود.

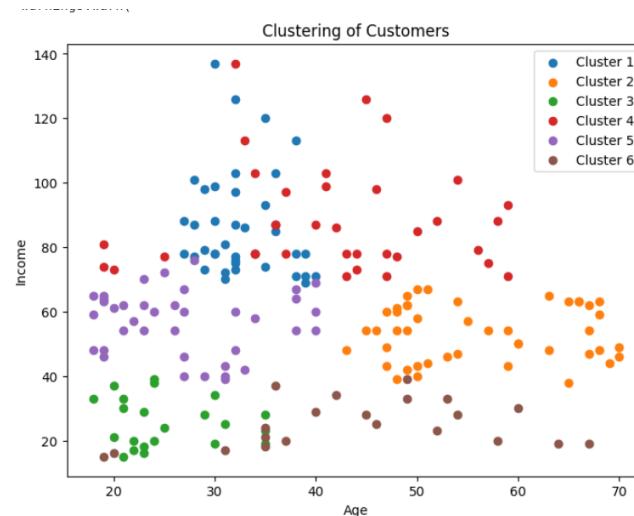
با توجه به نمودارهای فوق تعداد ۶ خوش، بهترین حالت برای خوشبندی خواهد بود.



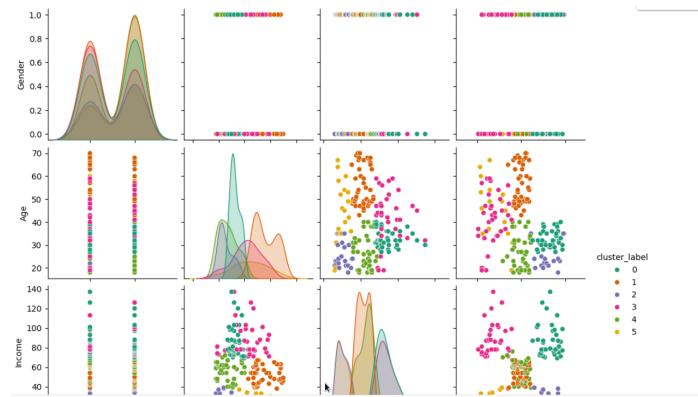
شکل ۱۷: انواع نمودارهای ذکر شده برای خوشبندی نقاط داده مشتریان

برای نمایش خوشبها در داده‌های بیش از ۲ بعدی ممکن است چالش‌هایی وجود داشته باشد، اما چند روش وجود دارد که می‌تواند به نمایش خوشبها به طور موثر کمک کند:

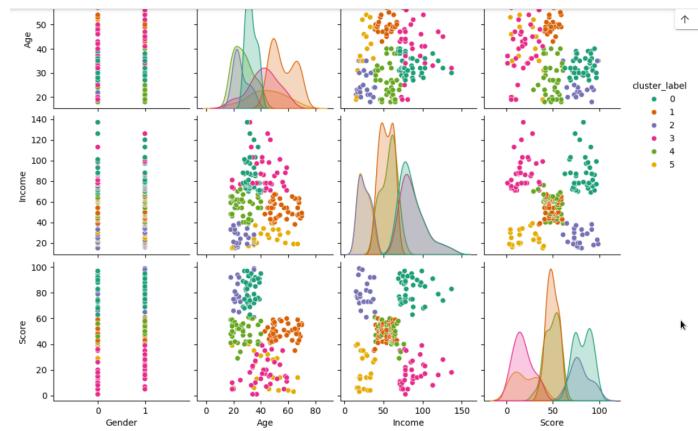
۱. تکنیک‌های کاهش ابعاد: یک رویکرد معمول استفاده از تکنیک‌های کاهش ابعاد مانند تجزیه و تحلیل مؤلفه‌های اصلی (PCA) یا توزیع t نزدیک همسایه تصادفی (t -SNE) است برای کاهش داده‌ای بعد بالا به یک فضای کم‌ابعاد (معمولاً ۲D یا ۳D) که به راحتی قابل نمایش است. پس از کاهش داده، می‌توانید خوشبها را با استفاده از نمودارهای پراکندگی سنتی یا نمودارهای پراکندگی سه‌بعدی نمایش دهید.
۲. نمودار مختصات موازی: یک روش دیگر استفاده از نمودار مختصات موازی است که یک تکنیک برای نمایش داده‌های بعد بالا با رسم هر نقطه داده به عنوان یک خط که از طریق مجموعه‌ای از محورهای موازی، یک برای هر بعد، عبور می‌کند. خوشبها با گروه‌بندی خطوط مشابه شناسایی می‌شوند. این روش می‌تواند به خوبی روابط بین ابعاد مختلف و نحوه مشارکت آن‌ها در خوشبندی را نشان دهد.



شکل ۱۸: شش خوشب برای داده مشتریان بر اساس دو ویژگی سن و حقوق دریافتی



شکل ۱۹: نمودارهای مختلف برای شش خوشه



شکل ۲۰: نمودارهای مختلف برای شش خوشه