گزارش تمرین پنجم یادگیری ماشین

بیتا آذری جو ۸۳۰۵۹۸۰۰۷

بیتا آذری جو — تمرین پنجم یادگیری ماشین — 830598007

سؤال اول:

i) $m_{n+1} = m_n + \frac{1}{n+1}(x_{n+1} - m_n)$

$$m_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{x_{n+1} + \sum_{i=1}^{n} x_i}{n+1} = \frac{x_{n+1} + n\,m_n}{n+1}$$

$$= m_n + \frac{x_{n+1} - m_n}{n+1} =$$

$$= m_n + \frac{1}{n+1}(x_{n+1} - m_n)$$

$$C_{n+1} = \frac{n-1}{n} C_n + \frac{1}{n+1}(x_{n+1} - m_n)(x_{n+1} - m_n)^T$$

طبق تعریف $C_{n+1} = \frac{1}{n}\sum_{i=1}^{n+1}(x_i - m_{n+1})(x_i - m_{n+1})^T$

$$C_n = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_n)(x_i - m_n)^T$$

می‌دانیم $\sum_{i=1}^{n}(x_i - m_n)(x_n - m_n)^T = \sum_{i=1}^{n} x_i x_i^T - n\,m_n m_n^T$

$$n C_{n+1} - (n-1) C_n = \sum_{i=1}^{n+1}(x_i - m_{n+1})(x_i - m_{n+1})^T - \sum_{i=1}^{n}(x_i - m_n)(x_i - m_n)^T$$

$$n C_{n+1} - (n-1) C_n = \sum_{i=1}^{n+1} x_i x_i^T - (n+1) m_{n+1} m_{n+1}^T - \sum_{i=1}^{n} x_i x_i^T + n\,m_n m_n^T$$

$$n C_{n+1} - (n-1) C_n = x_{n+1} x_{n+1}^T - (n+1)\left(\frac{n m_n + x_{n+1}}{n+1}\right)\left(\frac{n m_n + x_{n+1}}{n+1}\right)^T + n\,m_n m_n^T$$

$$= x_{n+1} x_{n+1}^T - \frac{(n+1)}{(n+1)^2}\left(n^2 m_n m_n^T + n\,m_n x_{n+1}^T + n\,x_{n+1} m_n^T + x_{n+1} x_{n+1}^T\right) + n\,m_n m_n^T$$

$$= x_{n+1} x_{n+1}^T - \frac{1}{n+1}\left(n^2 m_n m_n^T + n\,m_n x_{n+1}^T + n\,x_{n+1} m_n^T + x_{n+1} x_{n+1}^T\right) + n\,m_n m_n^T$$

$$= \frac{n}{n+1} x_{n+1} x_{n+1}^T - \frac{n}{n+1}\left(m_n x_{n+1}^T + x_{n+1} m_n^T\right) - \frac{n}{n+1} m_n m_n^T$$

$$n C_{n+1} - (n-1) C_n = \frac{n}{n+1}(x_{n+1} - m_n)(x_{n+1} - m_n)^T \Rightarrow$$

$$C_{n+1} = \frac{n-1}{n} C_n + \frac{1}{n+1}(x_{n+1} - m_n)(x_{n+1} - m_n)^T$$

$$C_{n+1} = \underbrace{\frac{n-1}{n} C_n}_{O(1)} + \underbrace{\frac{1}{n+1} (X_{n+1} - m_n)(X_{n+1} - m_n)^T}_{O(d^2)}$$

$O(d^3)$

(iii) زمان اجرا $O(d^3)$

(iii)

$$C_{n+1}^{-1} = \frac{n}{n+1} \left[C_n^{-1} - \frac{C_n^{-1}(X_{n+1} - m_n)(X_{n+1} - m_n)^t C_n^{-1}}{\frac{n^2-1}{n} + (X_{n+1} - m_n)^t C_n^{-1} (X_{n+1} - m_n)} \right]$$

$$C_{n+1}^{-1} = \left[\frac{n-1}{n} C_n + \frac{1}{n+1} (X_{n+1} - m_n)(X_{n+1} - m_n)^T \right]^{-1}$$

استفاده از نتیجه آمده شده در hint :

$$= \frac{n}{n-1} C_n^{-1} - \frac{\frac{n}{n-1} C_n^{-1} \frac{1}{n+1} (X_{n+1} - m_n)(X_{n+1} - m_n)^T \frac{n}{n-1} C_n^{-1}}{1 + \frac{1}{n+1} (X_{n+1} - m_n)^T \frac{n}{n-1} C_n^{-1} (X_{n+1} - m_n)} =$$

$$= \frac{n}{n-1} \left[C_n^{-1} - \frac{n C_n^{-1} (X_{n+1} - m_n)(X_{n+1} - m_n)^T C_n^{-1}}{\frac{n^2-1}{n} + (X_{n+1} - m_n)^T C_n^{-1} (X_{n+1} - m_n)} \right]$$

(iv) فرض می‌کنیم C_n^{-1} از مرحله قبل در حافظه داریم و نیازی به محاسبه مجدد ندارد. بنابراین:

$O(d^2)$ $O(d^2)$

$$\frac{n}{n-1} \left[C_n^{-1} - \frac{n C_n^{-1} (X_{n+1} - m_n)(X_{n+1} - m_n)^T C_n^{-1}}{\frac{n^2-1}{n} + \underbrace{(X_{n+1} - m_n)^T C_n^{-1} (X_{n+1} - m_n)}_{O(d^2)}} \right]$$

$O(d^2)$

محاسبات ـ روش بازگشتی سریع‌تر از روش عادی خواهد بود و به n بستگی ندارد.

سوال دوم:

۱. نتایج بر روی Iris dataset

i) نتایج به دست آمده به اینصورت شد که بدون PCA و FDA خطای کلی که با روش cross validation به دست آمده بود کم و منطقی بود. هنگام استفاده از PCA و کاهش به دو بعد دقت مقداری کاهش پیدا کرد. چون ممکن است اختلاف واریانس ویژگی ها کم باشد و ویژگی سوم یا چهارم واریانس نسبتا خوبی داشته باشند و با این کار که حذف شوند دقت مقداری کم می شود. وقتی به جای PCA، FDA با دو بعد اعمال شد. دقت نسبت به حالت اول حتی کمی بیشتر شد که نشان دهنده ی کارا بودن این روش می باشد چون تا حد ممکن داده ها را روی محوری map می کند که بیشترین جداپذیری را داشته باشند.

```
bita@bita-K401UQK:~/programming/Discriminant Analysis$ python LDA.py
Testing classifier on Iris...
Testing LDA classifier wihout PCA or FDA :
Cross Validation Missclassification error is :  17%
Cross Validation Accuracy is :  83%

Testing classifier with PCA :
Cross Validation Missclassification error is :  20%
Cross Validation Accuracy is :  80%

Testing classifier with FDA
Cross Validation Missclassification error is :  12%
Cross Validation Accuracy is :  88%
```

۲. نتایج بر روی Vowel dataset

کاهش بعد با روش FDA روی این داده موثرترین بوده چون توانسته بیشترین جدایی را ایجاد کند. PCA بدترین دقت را داد چون تعداد ویژگی ها از ۱۰ تا به دو تا کاهش یافت و ممکن است اختلاف واریانس بین سومین پراهمیت ترین ویژگی و دومین پراهمیت ترین ویژگی کم باشد. پس منطقی بود که feature های دیگر همین طوری حذف نشوند. در کل استفاده از LDA روی این دیتاست مناسب نبود و دقت کمی گرفتیم.

```
Testing classifier on vowel...
Testing on vowel dataset without PCA or FDA :
Missclassification error is :  43%
Accuracy is :  57%

Testing on vowel with PCA ...
Missclassification error is :  63%
Accuracy is :  37%

Testing on vowel with FDA ...
Missclassification error is :  35%
Accuracy is :  65%
```

منابع :

[1] https://xavierbourretsicotte.github.io/LDA_QDA.html

[2]https://towardsdatascience.com/gaussian-discriminant-analysis-an-example-of-generative-learning-algorithms-2e336ba7aa5c

[3] https://sebastianraschka.com/Articles/2014_pca_step_by_step.html

[4] https://sebastianraschka.com/Articles/2014_python_lda.html

[5] https://goelhardik.github.io/2016/10/04/fishers-lda/