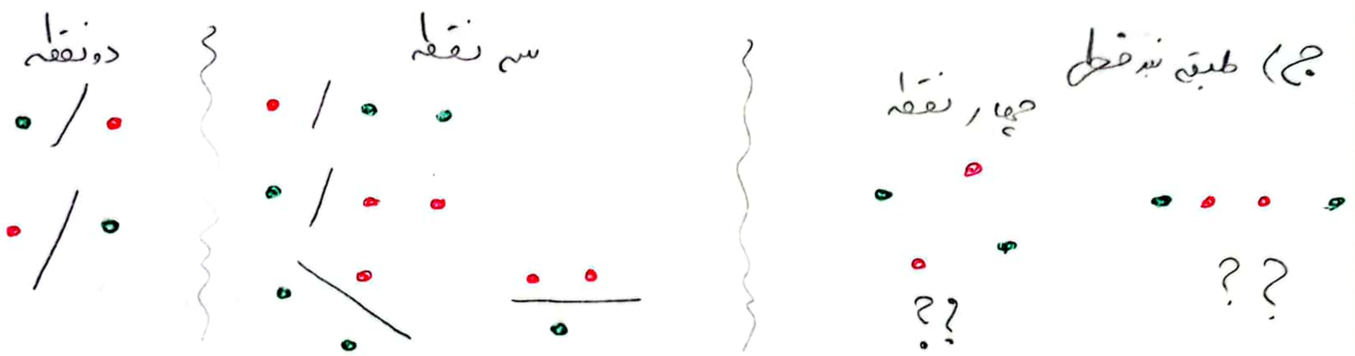
فاطمه ناشیان ٨١٠١٩٨٤٧٩ تمرین ۵ یادگیری ماشین

سوال ۱)

الف) vc dimension تعداد نقاط آموزشی است که می‌توان آن‌ها را shatter یا جدا کرد که این اندازه‌گیری میزان ظرفیت (پیچیدگی یا انعطاف‌پذیری) را نشان می‌دهد.

ب) vc dimension به عنوان مقدار توانایی و قدرتمندی یک فضا استفاده می‌شود که نشانگر بزرگترین مجموعه‌ای که می‌توان داده‌ها را shatter کرد که نشان دهنده این است که یک طبقه‌بند به خوبی بر روی آن مجموعه داده آموزش یابد

ج) طبقه بند خطی



بنابراین vc dimension برای یک طبقه‌بند خطی حداقل برابر ۳ و کمتر از ۴ است.

سوال ۲)

الف)

PCA: ماتریس covariance خیلی بزرگ می‌شود مثلاً برای تصاویر $N \times N$ ماتریس $N^2 \times N^2$ می‌شود. برای جلوگیری از این اتفاق ماتریس $M \times M$ را پیدا می‌کنیم که M تعداد جمله‌هاست سپس مقادیر ویژه ماتریس را حساب کرده و با کمک بردار ویژه‌ها مسیر را طی می‌کنیم.

LDA: اگر ابعاد زیادی داشته باشیم S_w تکین (singular) می‌شود $(S_w^{-1} S_b u_k = \lambda_k u_k)$. برای جلوگیری از این شکل ابتدا با کمک PCA بعدها را کم کرده و سپس LDA می‌زنیم.

ب)

$$X_1 = \left\{ \begin{pmatrix} 4.2 \\ 1.3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3.5 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 5.8 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix} \right\} \qquad \mu_1 = \left\{ \begin{pmatrix} 3.04 \\ 3.92 \end{pmatrix} \right\}$$

$$X_2 = \left\{ \begin{pmatrix} 8 \\ 10 \end{pmatrix}, \begin{pmatrix} 6 \\ 8.2 \end{pmatrix}, \begin{pmatrix} 9 \\ 5 \end{pmatrix}, \begin{pmatrix} 8 \\ 7 \end{pmatrix}, \begin{pmatrix} 10 \\ 9 \end{pmatrix} \right\} \qquad \mu_2 = \left\{ \begin{pmatrix} 8.2 \\ 7.84 \end{pmatrix} \right\}$$

$$S_1 = \sum_{x \in C_1} (x-\mu_1)(x-\mu_1)^T = \begin{bmatrix} 1.16 & -1.04 & -1.04 & 0.04 & 0.86 \\ -2.62 & -0.42 & 0.08 & 1.88 & 1.08 \end{bmatrix} \begin{bmatrix} 1.16 & -2.62 \\ -1.04 & -0.42 \\ -1.04 & 0.08 \\ 0.04 & 1.88 \\ 0.86 & 1.08 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 4.25 & -1.6816 \\ -1.6816 & 11.748 \end{bmatrix} =$$

$$S_2 = \sum_{x \in C_2} (x-\mu_2)(x-\mu_2)^T = \begin{bmatrix} -0.2 & -2.2 & 0.8 & -0.2 & 1.8 \\ 2.16 & 0.36 & -2.84 & 0.84 & 1.16 \end{bmatrix} \begin{bmatrix} -0.2 & 2.16 \\ -2.2 & 0.36 \\ 0.8 & -2.84 \\ -0.2 & 0.84 \\ 1.8 & 1.16 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 8.8 & -1.576 \\ -1.576 & 14.912 \end{bmatrix}$$

$$\star \; S_W = S_1 + S_2 = \begin{bmatrix} 13.05 & -3.2576 \\ -3.2576 & 26.66 \end{bmatrix}$$

$$\star \; S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^T = \begin{pmatrix} -5.16 \\ -3.92 \end{pmatrix} \begin{pmatrix} -5.16 & -3.92 \end{pmatrix} = \begin{pmatrix} 26.6256 & 20.2272 \\ 20.2272 & 26.6256 \end{pmatrix}$$

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

$$w^* = S_w^{-1}(\mu_1-\mu_2) = \begin{bmatrix} -0.4457 \\ -0.2015 \end{bmatrix} \qquad y = w^T x$$

سوال ٤) الف)

$$X_{+1} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} \qquad X_{-1} = \begin{bmatrix} 2.5 & 3 \\ 1 & 2 \end{bmatrix} \Longrightarrow \mu_{+1} = \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} \qquad \mu_{-1} = \begin{bmatrix} 2.75 \\ 1.5 \end{bmatrix}$$

$$S_B = (\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T = \begin{bmatrix} -1.25 \\ 1 \end{bmatrix} \begin{bmatrix} -1.25 & 1 \end{bmatrix} = \begin{bmatrix} 1.5625 & -1.25 \\ -1.25 & 1 \end{bmatrix}$$

ب)

$$S_1 = (x_{+1}-\mu_{+1})(x_{+1}-\mu_{+1})^T = \begin{bmatrix} -۰٫۵ & ۰٫۵ \\ -۱٫۵ & ۱٫۵ \end{bmatrix}\begin{bmatrix} -۰٫۵ & -۱٫۵ \\ ۰٫۵ & ۱٫۵ \end{bmatrix} = \begin{bmatrix} ۰٫۵ & ۱٫۵ \\ ۱٫۵ & ۴٫۵ \end{bmatrix}$$

$$S_{-1} = (x_{-1}-\mu_{-1})(x_{-1}-\mu_{-1})^T = \begin{bmatrix} -۰٫۲۵ & ۰٫۲۵ \\ -۰٫۵ & +۰٫۵ \end{bmatrix}\begin{bmatrix} -۰٫۲۵ & -۰٫۵ \\ ۰٫۲۵ & ۰٫۵ \end{bmatrix} = \begin{bmatrix} ۰٫۱۲۵ & ۰٫۲۵ \\ ۰٫۲۵ & ۰٫۵ \end{bmatrix}$$

$$S_w = S_1 + S_{-1} = \begin{bmatrix} ۰٫۶۲۵ & ۱٫۷۵ \\ ۱٫۷۵ & ۵ \end{bmatrix}$$

ج)

$$\omega^* = S_w^{-1}(\mu_1-\mu_2) = \begin{bmatrix} -۱۲۸ \\ ۴۵ \end{bmatrix} \rightarrow z = [-۱۲۸ \;\; ۴۵]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

x_i^T	y_i	z_i
(۱,۱)	۱	-۸۳
(۲,۴)	۱	-۷۶
(۲٫۵,۱)	-۱	-۲۷۵
(۳,۲)	-۱	-۲۹۴

سؤال ۵)

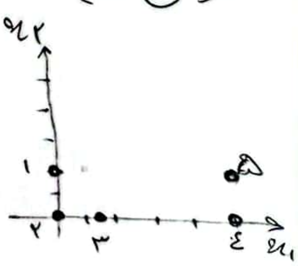
	x_1	x_2
x_1^T	۰	۲
x_2^T	۰	۰
x_3^T	۱٫۵	۰
x_4^T	۵	۰
x_5^T	۵	۲

$C_1 = \{x_1, x_2, x_4\}$

$C_2 = \{x_3, x_5\}$

$L_1\text{-norm} = \|x\|_1 = \sum |x_i|$

$L_2\text{-norm} = \|x\|_2 = \sqrt{\sum x_i^2}$



الف)

$$C_1 \rightarrow \mu_1 = \begin{bmatrix} ۵/۳ \\ ۲/۳ \end{bmatrix} \qquad C_2 \rightarrow \mu_2 = \begin{bmatrix} ۳٫۲۵ \\ ۱ \end{bmatrix}$$

$\|x_1-\mu_1\| = \sqrt{۴٫۵۵}$ $\quad \|x_2-\mu_1\| = \sqrt{۳٫۲۲}$ $\quad \|x_3-\mu_1\| = \sqrt{۰٫۴۷۲}$

$\|x_4-\mu_1\| = \sqrt{۱۱٫۵۵}$ $\quad \|x_5-\mu_1\| = \sqrt{۱۲٫۸۸}$

$\|x_1-\mu_2\| = \sqrt{۱۱٫۵۶۲۵}$ $\quad \|x_2-\mu_2\| = \sqrt{۱۱٫۵۶۲۵}$ $\quad \|x_3-\mu_2\| = \sqrt{۴٫۰۶۲۵}$

$\|x_4-\mu_2\| = \sqrt{۴٫۰۶۲۵}$ $\quad \|x_5-\mu_2\| = \sqrt{۴٫۰۶۲۵}$

$\|x_1-\mu_1\| = 1.42$ $\qquad \|x_2-\mu_1\| = 0.83$ $\qquad \|x_3-\mu_1\| = 1.2$

$\|x_4-\mu_1\| = 4.55$ $\qquad \|x_5-\mu_1\| = 4.69$

$\|x_1-\mu_2\| = 5.09$ $\qquad \|x_2-\mu_2\| = 5.09$ $\qquad \|x_3-\mu_2\| = 3.64$

$\|x_4-\mu_2\| = 1$ $\qquad \|x_5-\mu_2\| = 1$

$\longrightarrow C_1 = \{x_1, x_2, x_3\} \qquad C_2 = \{x_4, x_5\}$

با توجه به اینکه دسته‌ها دیگر تغییر نکرده پس خروجی نهایی نیز همین دسته‌ها خواهد بود

ب)

$C_1 \longrightarrow \begin{bmatrix} 5.3 \\ 2.3 \end{bmatrix} = \mu_1 \qquad C_2 \rightarrow \mu_2 = \begin{bmatrix} 3.25 \\ 1 \end{bmatrix}$

$\|x_1-\mu_1\| = 3$ $\qquad \|x_2-\mu_1\| = 2.3$ $\qquad \|x_3-\mu_1\| = 0.83$

$\|x_4-\mu_1\| = 4$ $\qquad \|x_5-\mu_1\| = 1.64$

$\|x_1-\mu_2\| = 4.25$ $\qquad \|x_2-\mu_2\| = 3.25$ $\qquad \|x_3-\mu_2\| = 2.25$

$\|x_4-\mu_2\| = 2.75$ $\qquad \|x_5-\mu_2\| = 2.75$

$\longrightarrow C_1 = \{x_1, x_2, x_3\} \rightarrow \mu_1 = \begin{bmatrix} 0.5 \\ 2.3 \end{bmatrix} \qquad C_2 = \{x_4, x_5\} \rightarrow \mu_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$

$\|x_1-\mu_1\| = 1.83$ $\qquad \|x_2-\mu_1\| = 1.16$ $\qquad \|x_3-\mu_1\| = 1.66$

$\|x_4-\mu_1\| = 5.16$ $\qquad \|x_5-\mu_1\| = 5.83$

$\|x_1-\mu_2\| = 6$ $\qquad \|x_2-\mu_2\| = 6$ $\qquad \|x_3-\mu_2\| = 4.5$

$\|x_4-\mu_2\| = 1$ $\qquad \|x_5-\mu_2\| = 1$

$\longrightarrow C_1 = \{x_1, x_2, x_3\} \qquad C_2 = \{x_4, x_5\}$

دیگر دسته‌ها تغییر نکرده خروجی نهایی نیز همین دسته خواهد بود.

CamScanner

سوال ۱۰)

$$P(x)=\frac{\lambda^{x} e^{-\lambda}}{x!} \qquad \sum_{j=1}^{m} a_{ij}=1 \qquad N_{ij}\sim P(a_{ij}\lambda_j)$$

likelihood

$$\rightarrow L(N_{ij})_{ij}(\lambda)=\prod_{i=1}^{n}\prod_{j=1}^{m} e^{-a_{ij}\lambda_j}\frac{(a_{ij}\lambda_j)^{N_{ij}}}{N_{ij}!} \rightarrow$$ (از استقلال نمونه ها)

$$\log L(N_{ij})_{ij}(\lambda)=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(-\lambda_j a_{ij}+N_{ij}\log(a_{ij}\lambda_j)-\log(N_{ij}!)\right)$$ نسبت به λ_j مشتق میگیریم و

$$\xrightarrow{\frac{d}{d\lambda_j}} \sum_{i=1}^{n}-a_{ij}+\frac{1}{\lambda_j}E[(N_{ij}|Y_i)]=0 \qquad , \quad N_{ij}|Y_i\sim Bin\left(Y_i,\frac{a_{ij}\lambda_j}{\sum_{k=1}^{m}a_{ik}\lambda_k}\right)$$

$$\Rightarrow \sum_{i=1}^{n}-a_{ij}+\frac{1}{\lambda_j}\frac{Y_i a_{ij}\lambda_j^{old}}{\sum_{k=1}^{m}a_{ik}\lambda_k^{old}}=0 \Rightarrow \boxed{\lambda_j=\frac{\lambda_j^{old}}{\sum_{i=1}^{n}a_{ij}}\cdot\sum_{i=1}^{n}\frac{Y_i a_{ij}}{\sum_{k=1}^{m}a_{ik}\lambda_k^{old}}}$$

سوال ۱)

الف) $P(x)=N(x|\mu,C)$

$$P(z)=N(z|0,I) \qquad P(x|z)=N(x|wz+\mu,\sigma^2 I)$$

$x=wz+\mu+\varepsilon \rightarrow$ خطا $P(\varepsilon)=N(\varepsilon|0,\sigma^2 I)$

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

$$E[x]=E[wz+\mu+\varepsilon]=wE[z]+E[\mu]+E[\varepsilon]=\mu$$

$$C=E[(x-\mu)(x-\mu)^T]=E[(wz+\varepsilon)(wz+\varepsilon)^T]=E[wzz^Tw+wz\varepsilon^T+z^Tw^T\varepsilon+\varepsilon\varepsilon^T]$$

$$=wE[zz^T]w^T+0+0+E[\varepsilon\varepsilon^T]=ww^T+\sigma^2 I$$

ب) (Bishop chapter 2)

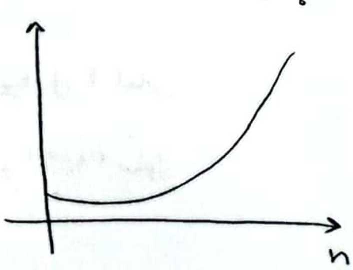
$$P(z|x)=N(z|M^{-1}w^T(x-\mu),\sigma^{-2}M)$$

$$C^{-1}=\sigma^2 I-\sigma^{-2}wM^{-1}w^T \rightarrow M=w^Tw+\sigma^2 I$$

سوال ۳) الف) $1\text{-}D: V=a \qquad 2\text{-}D: V=a^2 \qquad 3\text{-}D: V=a^3 \rightarrow n\text{-}D: V=a^n$

ب)

$$\left.\begin{aligned}&\text{hypercube}: V_a=a^n\\&\text{hypersphere}: V_r=\frac{\pi^{\frac{n}{2}}r^n}{\Gamma(\frac{n}{2}+1)}\end{aligned}\right\}\quad \frac{V_a}{V_r}=\Gamma\left(\frac{n}{2}+1\right)\left(\frac{2}{\pi}\right)^{\frac{n}{2}}$$



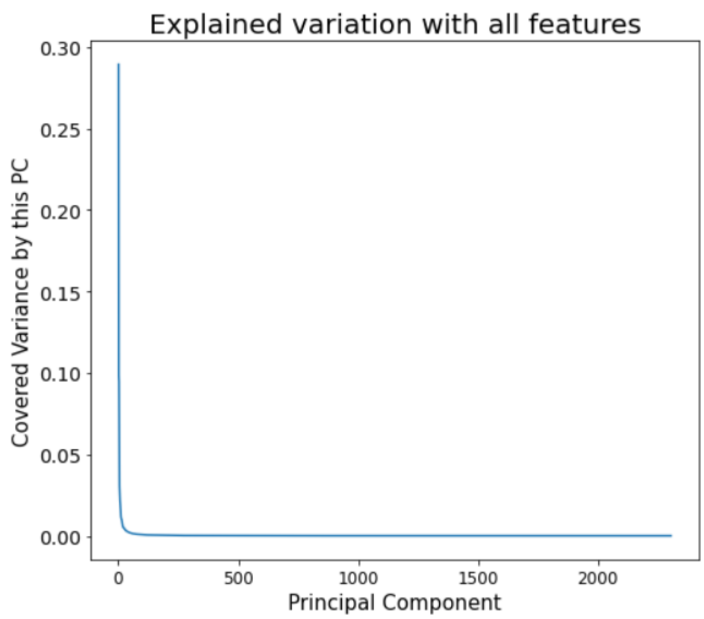
$$n=1\Rightarrow\frac{V_a}{V_r}=\frac{1}{\sqrt{2}} \qquad n=2\Rightarrow\frac{V_a}{V_r}=\frac{2}{\pi}$$

$$n=3\Rightarrow\frac{V_a}{V_r}=\frac{3}{\sqrt{2}\pi} \qquad n\Rightarrow\infty\Rightarrow\frac{V_a}{V_r}=\lim_{n\to\infty}\Gamma\left(\frac{n}{2}+1\right)\left(\frac{2}{\pi}\right)^{\frac{n}{2}}=\infty$$

6)

با توجه به نمودار زیر میبینیم بیشترین تاثیر را چند مقدار ویژه اول دارند که اندازه انها برابر است با :

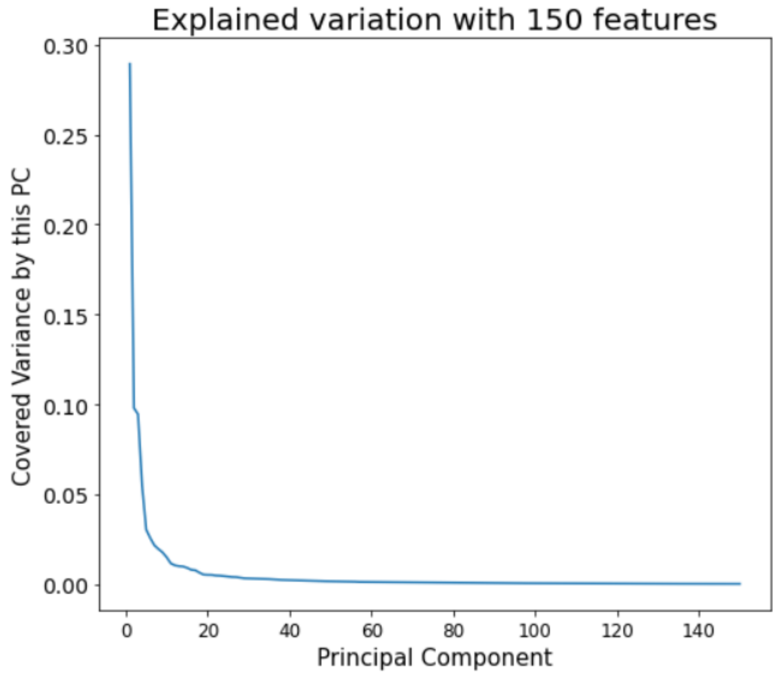
```
Explained variation per principal component for the first four PCs: [0.28923892 0.09806088 0.09456227 0.05485086]
```



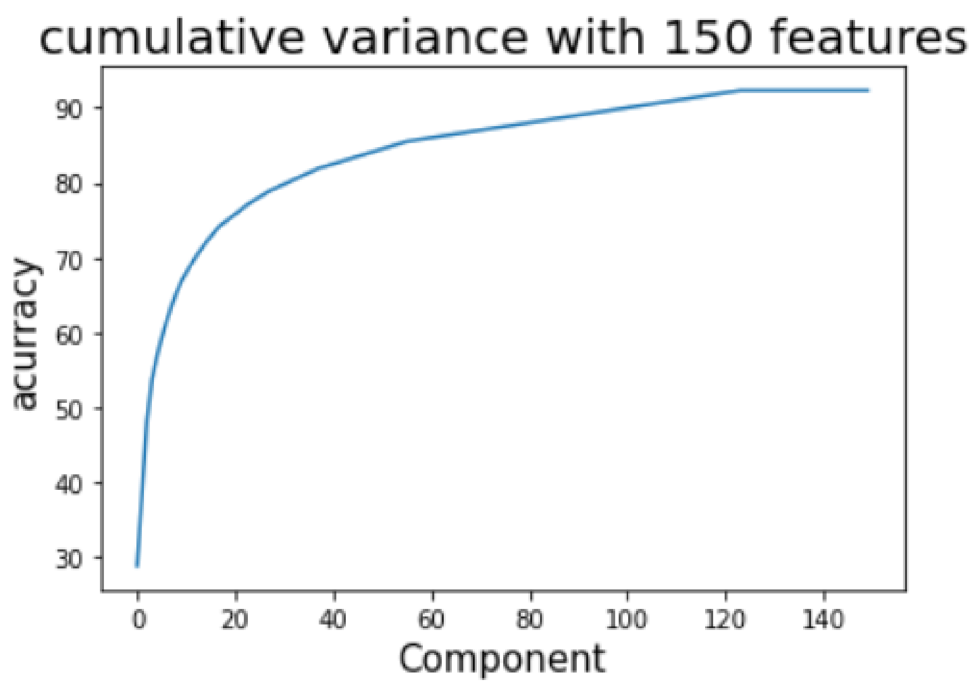
در شکل زیر نمودار واریانس تجمعی را میبینیم که نشان میدهد با تعداد فیچر های حدود 150 تا میتوانیم به دقت بالای 90 برسیم.

پس با کمک واریانس تجمعی میتوان تعداد کامپوننت مناسب را پیدا کرد

حال دوباره pca میزنیم و این بار فقط 150 تا فیچر را در نظر میگیریم.



واریانس تجمعی نیز نشان از دقت بالای 90 درصد می دهد.

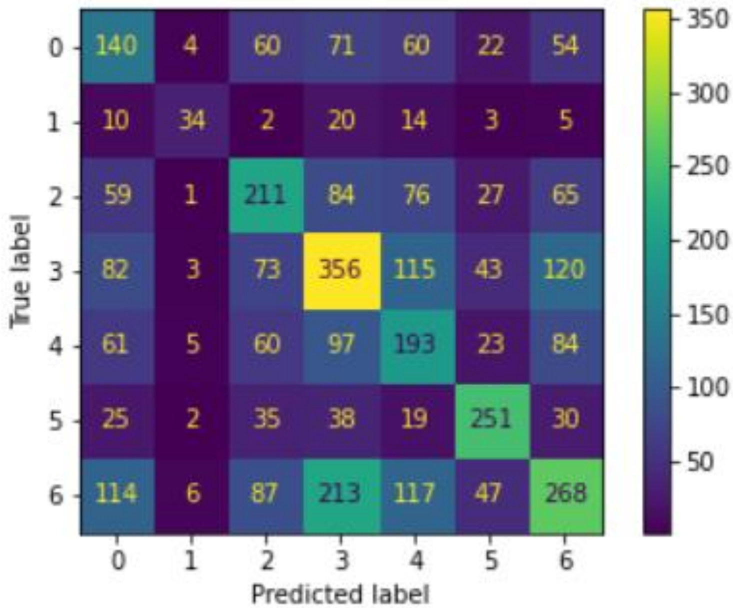


Eigenface های اول فراوانی بیشتری دارند.

حال knn میزنیم

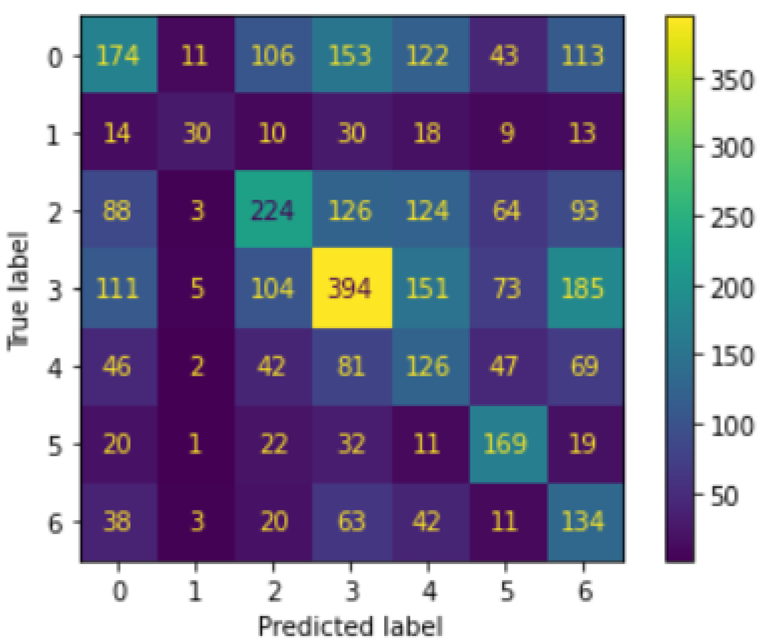
Knn با k=1 برای دادهای خالص

```
accuracy score on test data for : 0.40484814711618833
```



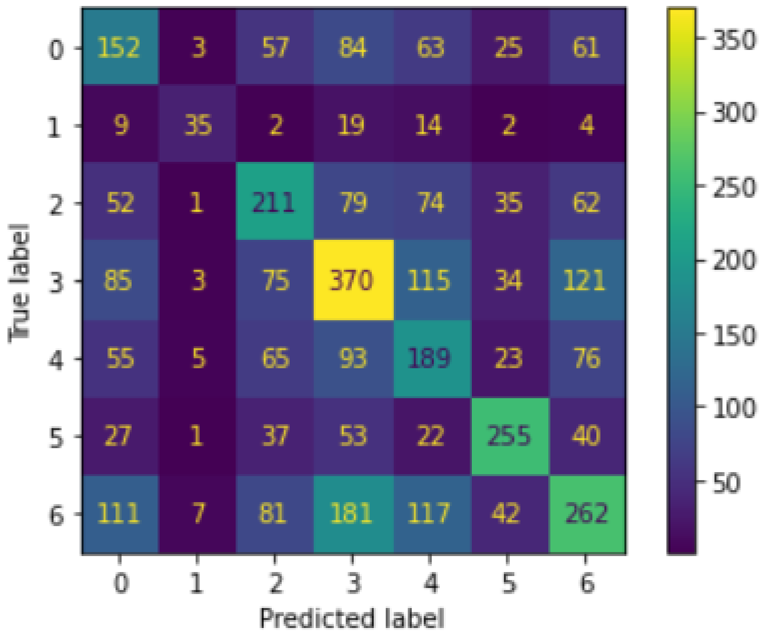
Knn با k=2 برای دادهای خالص

```
accuracy score on test data for : 0.3485650599052661
```



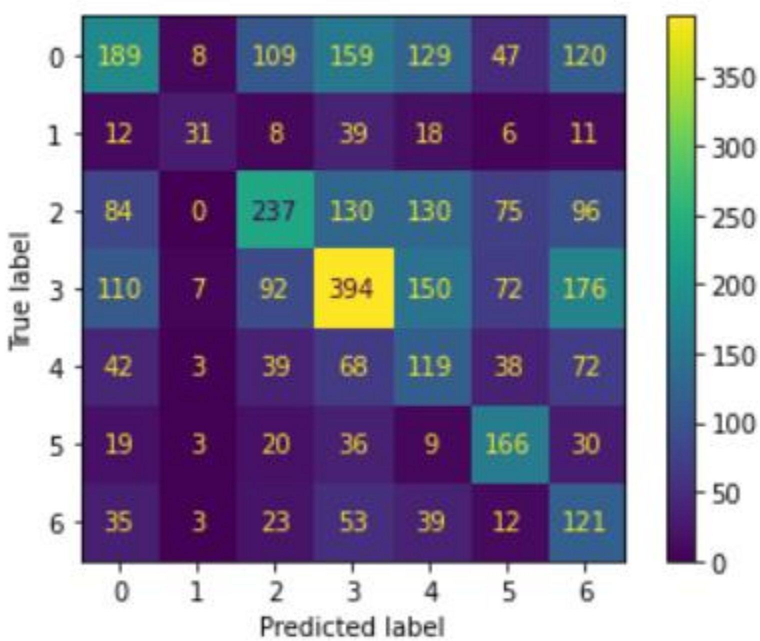
Knn با k=1 برای دادهای کاهش بعد یافته

```
accuracy score on test data for : 0.4106993591529674
```



Knn با k=2 برای دادهای کاهش بعد یافته

```
accuracy score on test data for : 0.3502368347729172
```



برای داده های کاهش بعد یافته با تعداد 1 همسایه بیشترین دقت و کمترین خطا را داریم.

```
for n in range(1,10):
    clf_knn = KNeighborsClassifier(n_neighbors=n)
    clf_knn.fit(new_x_train , y_train)
    pred = clf_knn.predict(new_x_test)
    score = accuracy_score(pred, y_test)
    print("accuracy score on test data with n={} for : {}".format(n ,score))
```

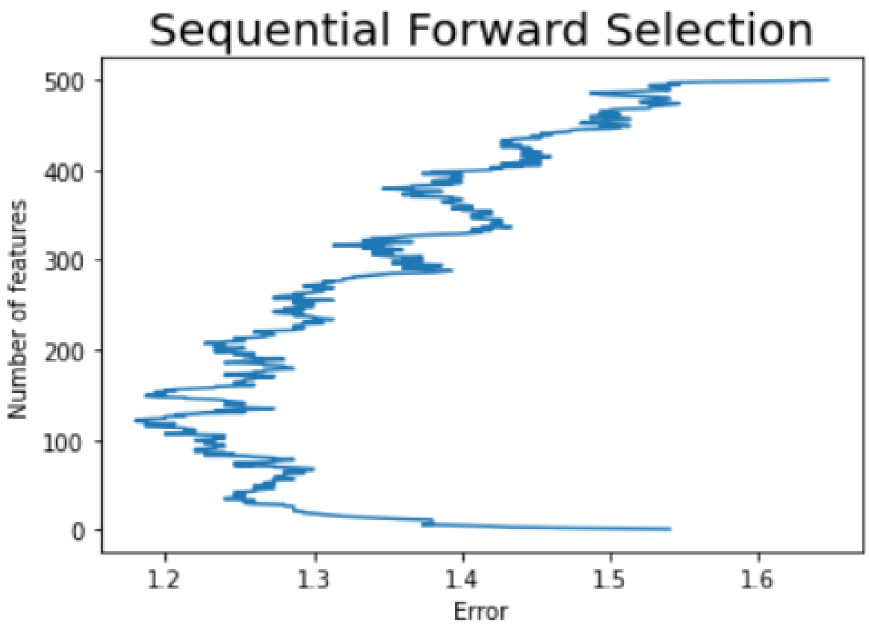
```
accuracy score on test data with n=1 for : 0.40484814711618833
accuracy score on test data with n=2 for : 0.3485650599052661
accuracy score on test data with n=3 for : 0.3390916689885762
accuracy score on test data with n=4 for : 0.3337977152410142
accuracy score on test data with n=5 for : 0.3365840066870995
accuracy score on test data with n=6 for : 0.32711061577040956
accuracy score on test data with n=7 for : 0.32683198662580104
accuracy score on test data with n=8 for : 0.32460295346893286
accuracy score on test data with n=9 for : 0.32850376149345223
```

(7

Sequential Forward Selection

نمودار خطا بر حسب تعداد ویژگی به صورت زیر رسم می شود.

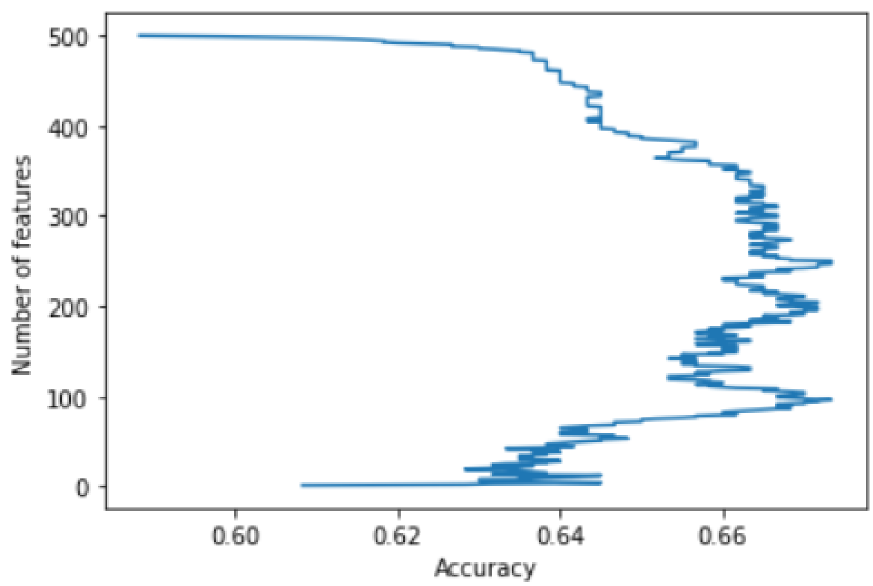
بر اساس این نمودار در اخر 120 ویژگی میتواند کمترین خطا را به ما بدهد.



Sequential Backward Elimination

نمودار دقت بر حسب تعداد ویژگی به صورت زیر رسم می شود.

بر اساس این نمودار در اخر 250 ویژگی میتواند کمترین خطا و بیشترین دقت را به ما بدهد.



از مزیت های backward میتوان به این اشاره کرد که سرعت و دقت این الگوریتم بیشتر است و خطای کمتری دارد.

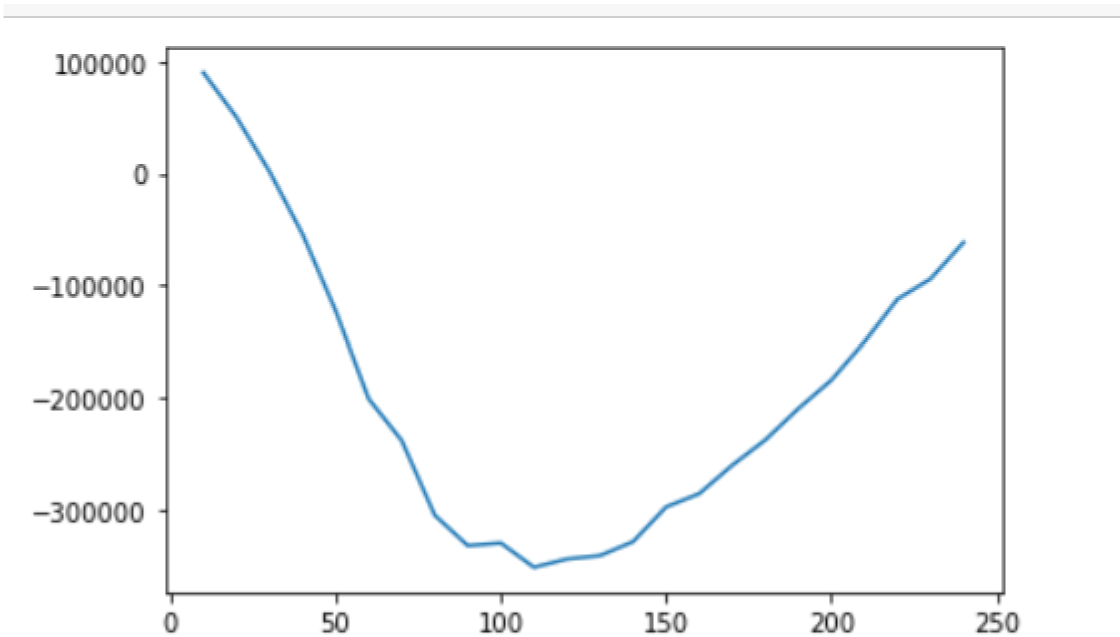
در forward فرایند به گونه ایست که هربار یک ویژگی اضافه می شود به صورتی که بیشترین دقت را در پیش بینی کردن ایجاد کند و فرایند انقدر ادامه پیدا میکند که ویژگی جدید در پیش بینی تاثیر مثبتی نداشته باشد.

در backward ابتدا همه ویژگی ها استفاده می شود و در هر مرحله ویژگی که تاثیر خاصی در پیش بینی نداشته باشد حذف می شود.

در کل backward بهتر کار میکند چون در forward در هر مرحله همه ی ویژگی های باقی مانده صدا می شوند و هزینه خیلی بیشتری دارد.

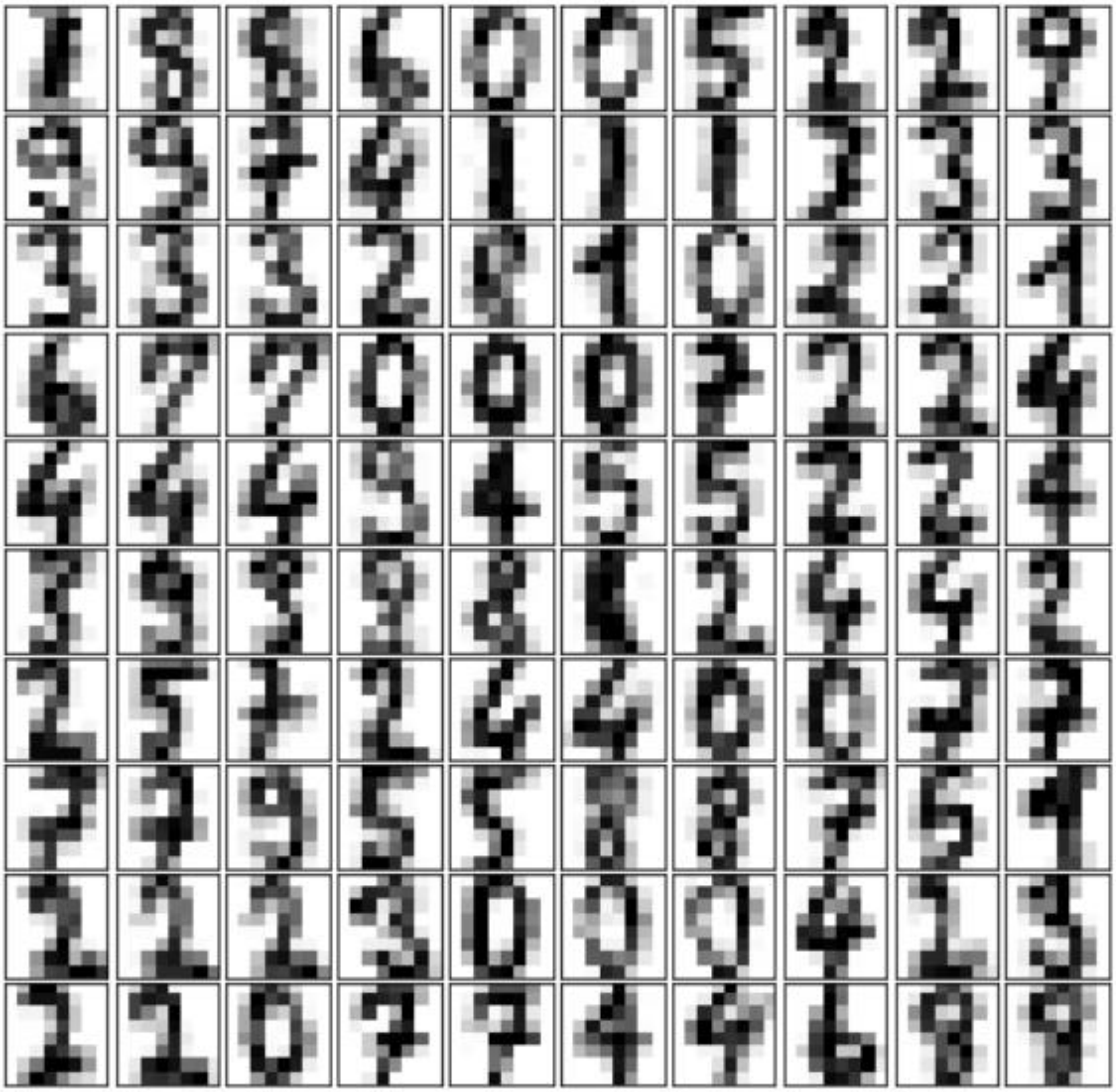
(8

نمودار aic را برای داده ها رسم میکنیم.



بهترین تعداد component ها 110 تا است.

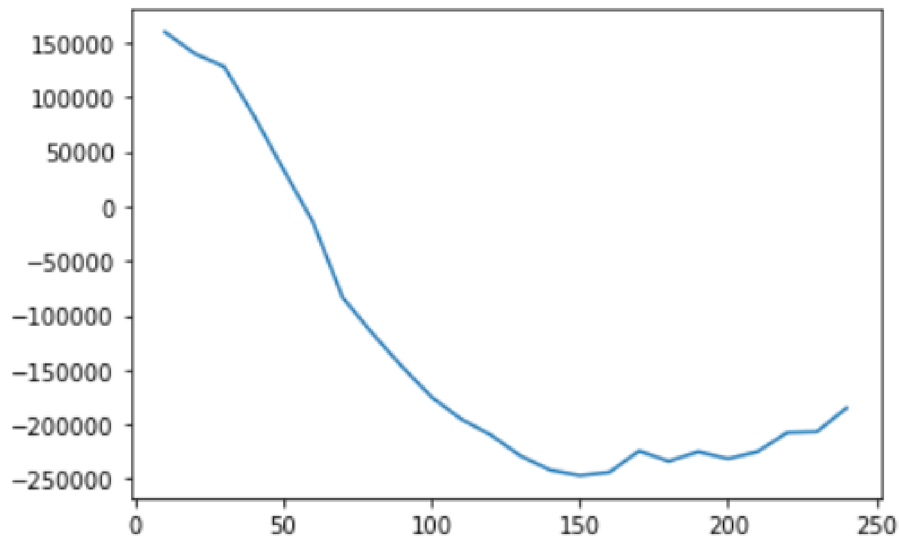
حال یک مدل با تعداد component=110 فیت میکنیم و از ان 100 سمپل میگیریم و نمایش میدهیم.

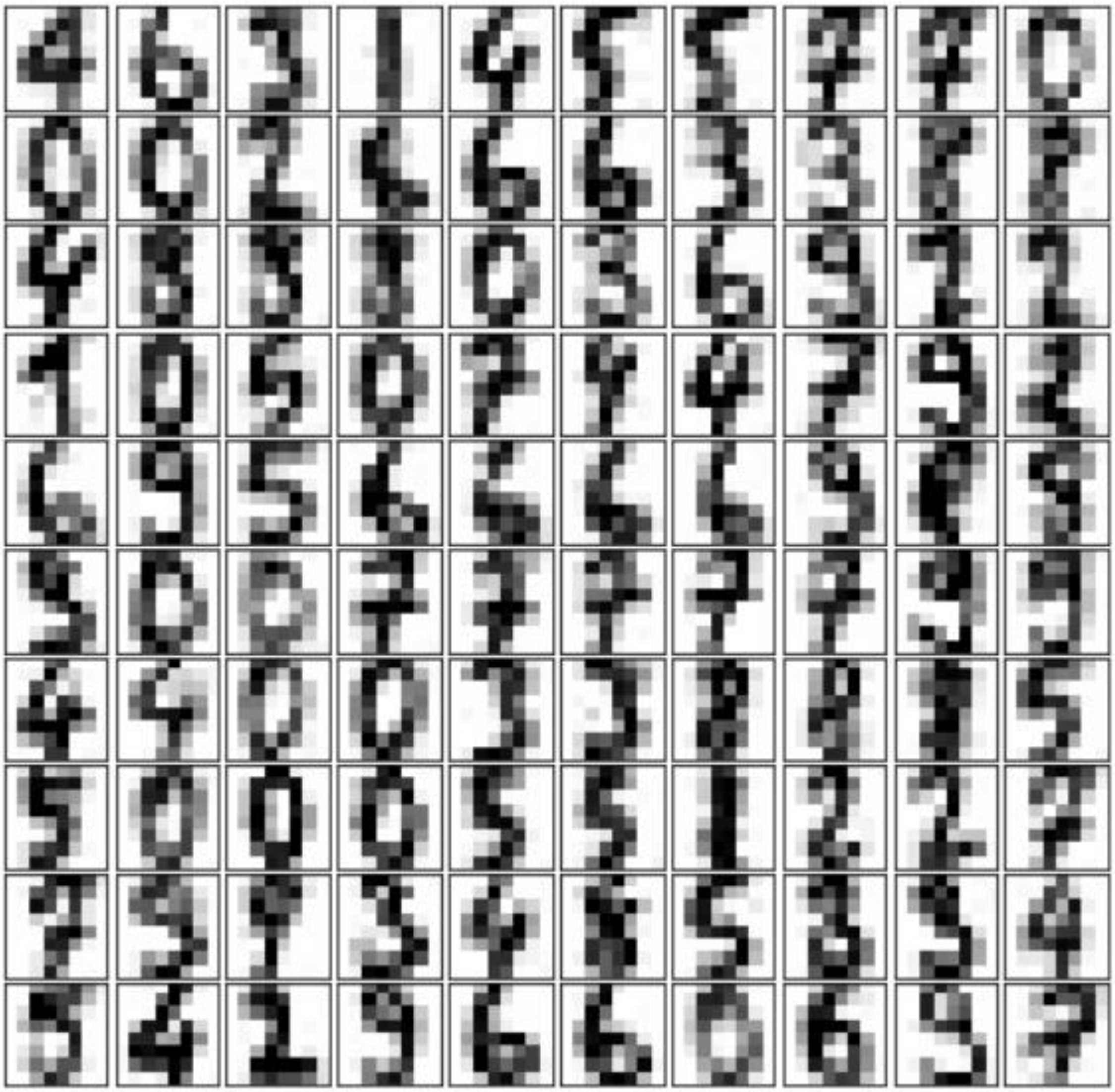


همگرایی رخ داده است.

حال ابتدا کاهش بعد میدهیم و سپس aic ان را نمایش میدهیم.

برای component های 150 تا کمترین aic را داریم. با این تعداد یک مدل فیت میکنیم.



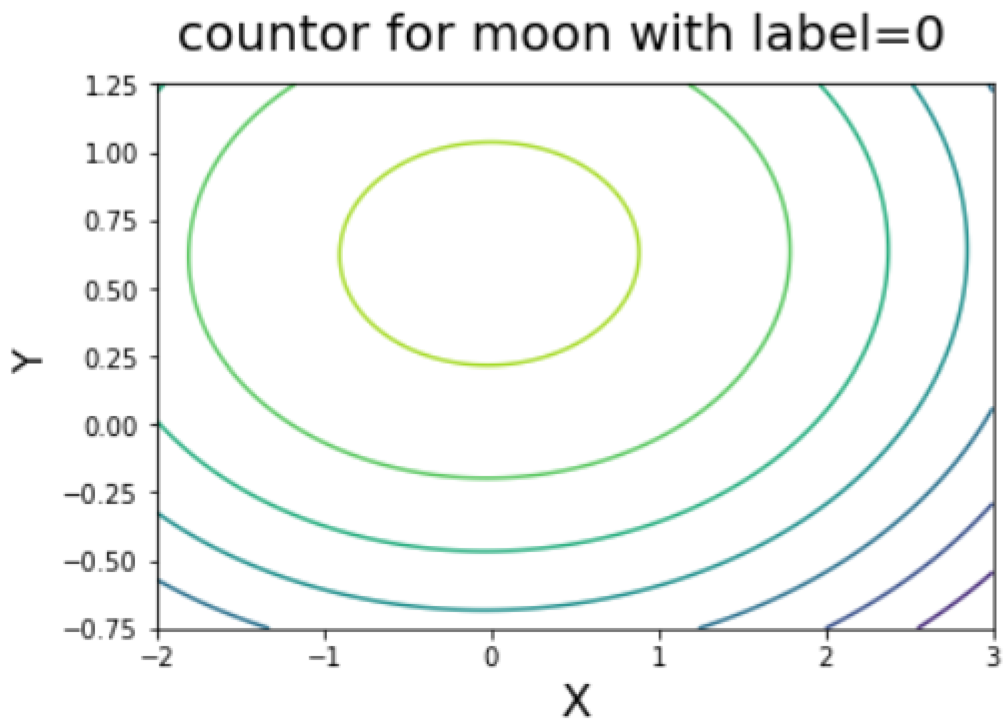


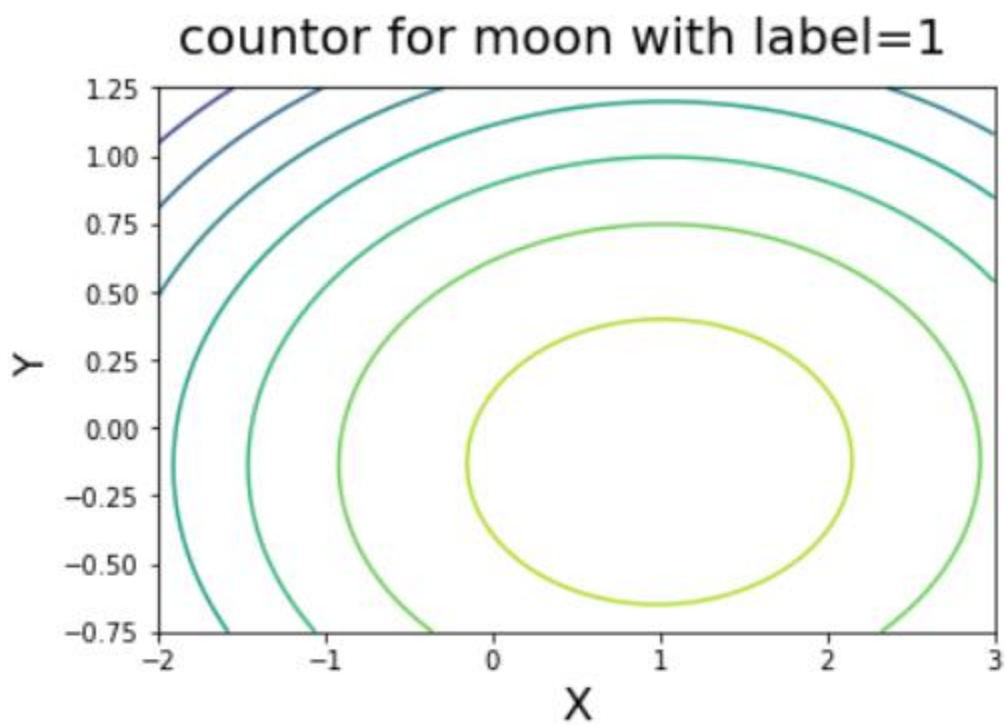
همگرایی رخ داده است.

نتایج در بخش ج بهتر شده است زیرا با pca تعداد ابعاد را کم کرده ایم در نتیجه کیفیت و نتایج بهتری داریم و دقت بالاتر می رود.

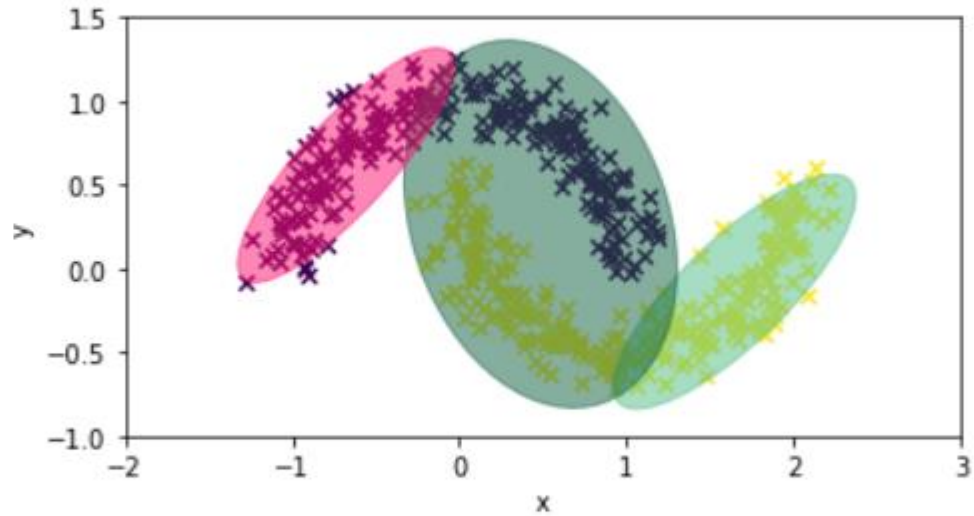
9)

ابتدا دونرمال فیت میکنیم و کانتور های انها را رسم میکنیم.

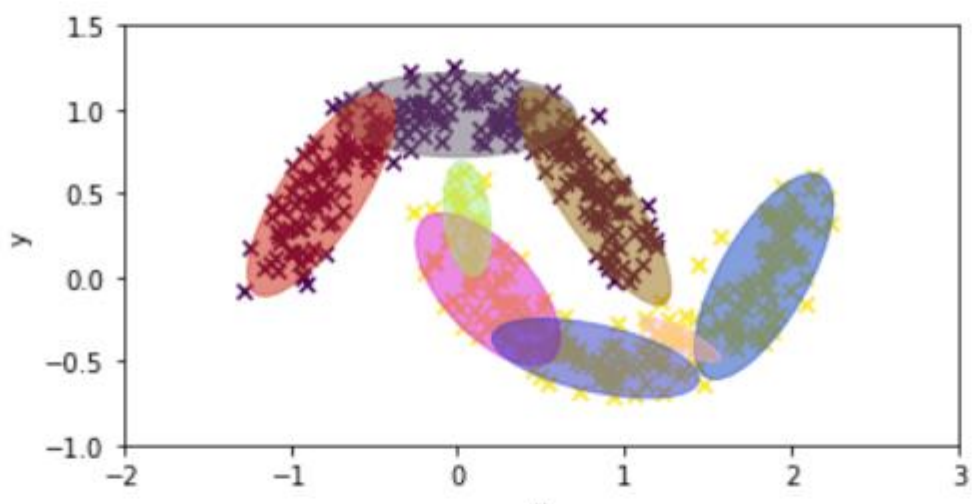




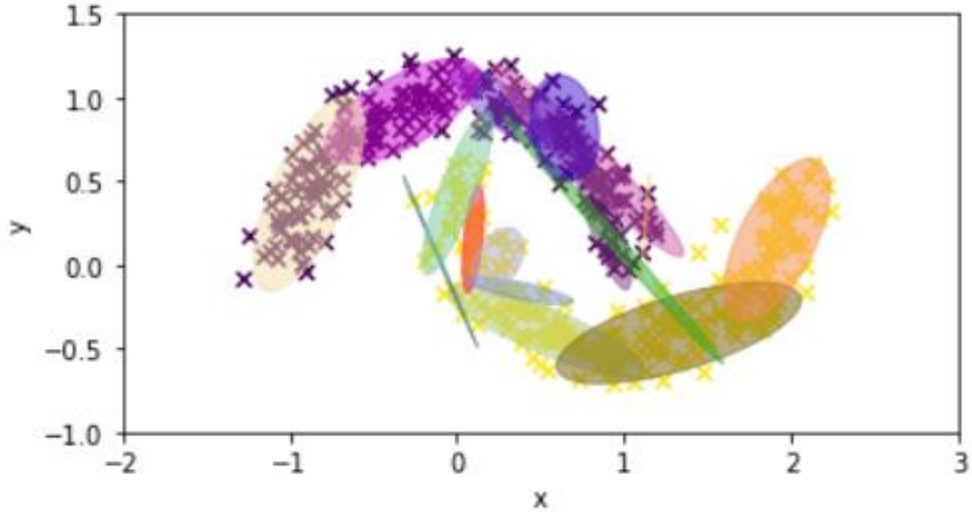
مولفه 3



مولفه 8



مولفه 16


1.5
1.0
0.5
0.0
−0.5
−1.0
y
−2
−1
0
1
2
3
x