به نام خدا

دانشکده مهندسی برق

گزارش درس یادگیری ماشین

مقطع: کارشناسی ارشد   گرایش: مهندسی کنترل

گزارش آزمون پایان ترم

توسط:

مرجان محمدی

۴۰۱۱۱۵۳۴

استاد درس:

دکتر علیاری

لینک کولب

لینک دوم کولب

لینک کولب

لینک گیت هاب

پرسش یک

آ. (هماهنگ‌شدهٔ ۱): در یک سیستمی که تفکیک پذیر خطی نیست، ثابت کنید اگر بخواهیم با SVM آن را کلاس‌بندی کنیم در تابع هزینه آن شرط زیر برقرار است :

(۱)

$$L(w, b, \Lambda, R, \varepsilon) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i \left(w^T x_i + b\right) - 1 + \varepsilon_i\right) - \sum_{i=1}^{N} r_i \varepsilon_i + C \sum \varepsilon_i^2$$

$$0 \le \alpha_i \le C$$

$$\Phi = \frac{1}{2} w.w + C\left(\sum_{i=1}^{N} \varepsilon_i\right)^2$$

شروط $\Rightarrow y_i(x_i w + b) \geq 1 - \varepsilon_i \quad i = 1, \dots, N \quad , \varepsilon_i \geq 0 \quad i = 1, \dots, N$

$$L(w, \varepsilon, b, \alpha, r) = \frac{1}{2} w w + C\left(\sum_{i=1}^{N} \varepsilon_i\right)^2 - \sum_{i=1}^{N} \alpha_i \left(y_i(x_i w + b) - 1 + \varepsilon_i\right) - \sum_{i=1}^{N} r_i \varepsilon_i$$

$\xrightarrow{1} r_i \geq 0 \; , \alpha_i \geq 0 \Rightarrow$ ضرایب لاگرانژ

$$\Rightarrow \frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \varepsilon_i} = 2C \sum_{i=1}^{N} \varepsilon_i - \alpha_i - r_i = 0 \Rightarrow \alpha_i + r_i = 2C \sum_{i=1}^{N} \varepsilon_i = \delta \Rightarrow \sum_{i=1}^{N} \varepsilon_i = \frac{\delta}{2C}$$

بافرض : $r_i \varepsilon_i = 0 \; , \; \varepsilon_i \geq 0 \Rightarrow \alpha_i = 2C \sum_{i=1}^{N} \varepsilon_i$

با جایگذاری w در L داریم:

$$L(\alpha, \varepsilon, b) = \frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)\cdot\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right) + C\left(\sum_{i=1}^{N} \varepsilon_i\right)^2 - \sum_{i=1}^{N} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} r_i \varepsilon_i$$

چون $w = \sum_{i=1}^{N} \alpha_i y_i x_i$ و b در ترم‌های باقیمانده ظاهر شده و $\sum_{i=1}^{N} \alpha_i y_i = 0$

$\Leftarrow$ L را به صورت زیر می‌نویسیم:

$$L(\alpha, \varepsilon) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + C\left(\sum_{i=1}^{N} \varepsilon_i\right)^2 - \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \overset{0}{\cancel{r_i}} \varepsilon_i$$

فرم دوگان این تابع هزینه با توجه به مقاله رفرنس به صورت زیر نوشته می‌شود:

$$W(\alpha, \delta) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \frac{\delta^2}{2C}\left(1 - \frac{1}{2}\right)$$

با توجه به اینکه $\alpha_i \geq 0$ و $\Lambda = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}$ و $D = \mathrm{diag}(x_i x_j y_i y_j)$

$$\Rightarrow W(\Lambda, \delta) = \Lambda^T 1 - \left(\frac{1}{2} \Lambda^T D \Lambda + \frac{\delta^2}{2C}\left(1 - \frac{1}{2}\right)\right) \quad (I)$$

با شرط‌های: $R \geq 0$, $\Lambda \geq 0$, $\Lambda + R = \delta 1$, $\Lambda^T Y = 0$

با توجه به $\Lambda + R = \delta 1$ و $R \geq 0$ $\Leftarrow$ $\Lambda = \delta 1 - R \rightarrow \Lambda \leq \delta 1$ و طبق: $\Lambda \geq 0$ پس:

$$0 \leq \Lambda \leq \delta 1$$

با فرض $\delta = \alpha_{max} = \max(\alpha_1, \ldots, \alpha_N)$ و با جایگذاری در تابع هزینه (I):

$$W(\Lambda) = \Lambda^T 1 - \left[\frac{1}{2} \Lambda^T D \Lambda + \frac{\alpha_{max}^2}{2C}\left(1 - \frac{1}{2}\right)\right]$$

با توجه به متن مقاله‌ی رفرنس QP , $k=2$ در نظر گرفته شود $\Leftarrow$ به شکل زیر بدست می‌آید:

$$W(\Lambda) = \Lambda^T 1 - \frac{1}{2} \Lambda^T D \Lambda$$

با شرط‌های زیر:

$$0 \leq \Lambda \leq C1 \ , \ \Lambda^T Y = 0$$

از قید اول داریم:

$$0 \le \lambda \le C1 \rightarrow 0 \le \alpha_i \le C$$

با این قید α بیشینه است و در نتیجه L نیز بیشینه است. پس ثابت می‌شود که قید $0 \le \alpha_i \le C$ تابع هزینه را بیشینه می‌کند.

ب. (هماهنگ‌شدهٔ ۲): ثابت کنید در روش کاهش بعد با LDA ، مقدار Scatter Total داده‌ها طبق رابطه زیر بدست می‌آید:

$$S_{\text{Total}} = S_B + S_w$$
$$S_B = \text{ Between Scatter } , S_w = \text{ Within Scatter} \tag{۲}$$

$$S_w = \sum_{i=1}^{c} S_i \ , \ S_i = \sum_{j=1}^{N_i} (x_i - \mu_i)(x_i - \mu_i)^T \Rightarrow \mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j$$

$$\mu = \frac{1}{n} \sum_{x} x = \frac{1}{n} \sum_{i=1}^{c} n_i \mu_i$$

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \ , \ S_T = \sum_{x} (x - \mu)(x - \mu)^T$$

$$\Rightarrow S_T = \sum_{x} (x - \mu)(x - \mu)^T$$

$$\sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_j - \mu_i + \mu_i - \mu)(x_j - \mu_i + \mu_i - \mu)^T = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_j - \mu_i)(x_j - \mu_i)^T + \sum_{i=1}^{c} \sum_{j=1}^{N_i} (\mu_i - \mu)(\mu_i - \mu)^T \quad \text{I}$$

$$\Rightarrow \sum_{i=1}^{c} \sum_{j=1}^{N_i} (\mu_i - \mu)(\mu_i - \mu)^T = \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T \sum_{j=1}^{N_i} 1 = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

$$\Rightarrow \text{I} = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_j - \mu_i)(x_j - \mu_i)^T + \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T = \boxed{S_w + S_B = S_T}$$

ج. در فضای دوبُعدی نقاط زیر و برچسب متناظر با آن‌ها داده شده است:

$\mathrm{x}_1 = \{(1,1),1\}, \quad \mathrm{x}_2 = \{(2,1),1\}, \quad \mathrm{x}_3 = \{(2,0),1\}, \quad \mathrm{x}_4 = \{(1,2),-1\}, \quad \mathrm{x}_5 = \{(2,2),-1\}, \quad \mathrm{x}_6 = \{(1,-3),-1\}$

چنان‌چه از تبدیل غیرخطی $\varphi(\mathrm{x}) = (x_1^2 + x_2^2, x_1 - x_2)$ برای رسیدن به یک ماشین بردار پشتیبان استفاده شود به سوالات زیر پاسخ دهید:

آ) بردارهای $\varphi_i(\mathrm{x}), i = 1, \ldots, 6$ را در فضای مخفی رسم کنید و بدون محاسبه و صرفاً به‌صورت شهودی بردارهای پشتیبان را پیدا کنید.

ب) با استفاده از حلّ مسألهٔ دوگان، مقادیر بهینهٔ وزن و بایاس را پیدا کنید.

ج) یک تبدیل غیرخطی معرفی کنید که داده‌ها را از فضای دوبُعدی به فضای یک‌بُعدی انتقال دهد، به‌طوری‌که دسته‌بندی خطی آن‌ها امکان‌پذیر شود.

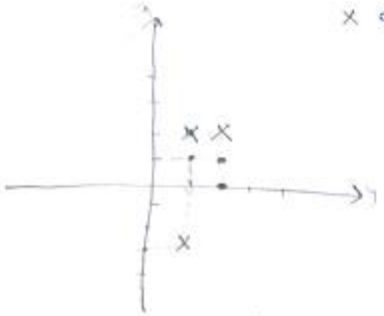
$x_1 = \{(1,1),1\} \quad x_2 = \{(2,1),1\} \quad x_3 = \{(2,0),1\} \quad x_4 = \{(1,2),-1\}$

$x_5 = \{(2,2),-1\} \quad x_6 = \{(1,-3),-1\}$

$\varphi(x) = (x_1^2 + x_2^2, x_1 - x_2)$

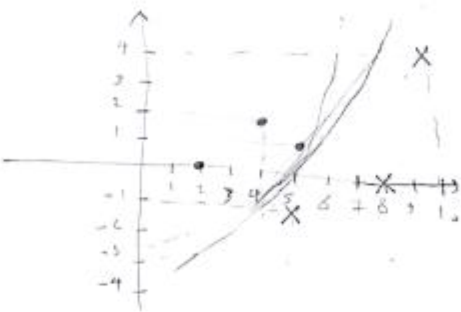
نقاط با برچسب $+1$: x_1 , x_2 , x_3 .

نقاط با برچسب -1 : x_4 , x_5 , x_6 ×



$\varphi_1(x) = (2,0) \qquad \varphi_2(x) = (5,1) \qquad \varphi_3(x) = (4,2)$

$\varphi_4 = (5,-1) \qquad \varphi_5 = (8,0) \qquad \varphi_6 = (10,4)$



بردارهای پشتیبان

$\varphi_4(x)$, $\varphi_2(x)$

دو α مربوط به φ_2 و φ_4 داریم که غیر صفر هستند، بقیه α ها صفر هستند

$$\alpha_2 \varphi_2^2 + \alpha_4 \varphi_2 \varphi_4 = 1$$

$$\alpha_2 \varphi_2 \varphi_4 + \alpha_4 \varphi_4^2 = -1$$

$$\Rightarrow \quad \alpha_2 \varphi_2 \varphi_2 + \alpha_4 \varphi_2 \varphi_4 = 1$$

$$\alpha_2 \varphi_2 \varphi_4 + \alpha_4 \varphi_4 \varphi_4 = -1$$

$$\alpha_2 \begin{pmatrix} 5 \\ 1 \end{pmatrix}^T \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 5 \\ 1 \end{pmatrix}^T \begin{pmatrix} 5 \\ -1 \end{pmatrix} = 1$$

$$\alpha_2 \begin{pmatrix} 5 \\ -1 \end{pmatrix}^T \begin{pmatrix} 5 \\ +1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 5 \\ -1 \end{pmatrix} \begin{pmatrix} 5 \\ -1 \end{pmatrix} = -1$$

$$\Rightarrow \quad 10\alpha_2 + 24\alpha_4 = 1$$

$$24\alpha_2 + 26\alpha_4 = -1$$

$$\Rightarrow \begin{cases} 240\alpha_2 + 576\alpha_4 = 24 \\ -240\alpha_2 - 260\alpha_4 = +10 \end{cases} \Rightarrow (576 - 260)\alpha_4 = 34$$

$$\Rightarrow 316\alpha_4 = 34 \Rightarrow \alpha_4 = \frac{34}{316} \simeq \boxed{0.1076} \approx 0.11$$

$$\Rightarrow \alpha_2 = \frac{1 - 24\alpha_4}{10} = \frac{1 - 2.58}{10} = \boxed{-0.158}$$

$$10 \qquad 10$$

$$\Rightarrow w_{optimal} = \alpha_2 y_2 \varphi_2 + \alpha_4 y_4 \varphi_4 = -0.158 \times +1 \times \begin{pmatrix} 5 \\ 1 \end{pmatrix} + 0.11 \times -1 \times \begin{pmatrix} 5 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} -0.079 \\ -0.158 \end{pmatrix} + \begin{pmatrix} -0.55 \\ 0.11 \end{pmatrix} = \begin{pmatrix} -0.629 \\ -0.048 \end{pmatrix}$$

$$b_{optimal} = \alpha_2 (y_2 \varphi_2^T + b) - 1 = 0 \Rightarrow \alpha_2 y_2 \varphi_2^T + \alpha_2 b = 1$$

$$\Rightarrow -0.158 \times 1 \times (5\ 1)\begin{pmatrix} 5 \\ 1 \end{pmatrix} + (0.158\, b) = 1 \Rightarrow -0.158 \times 26 - 1 = 0.158 b$$

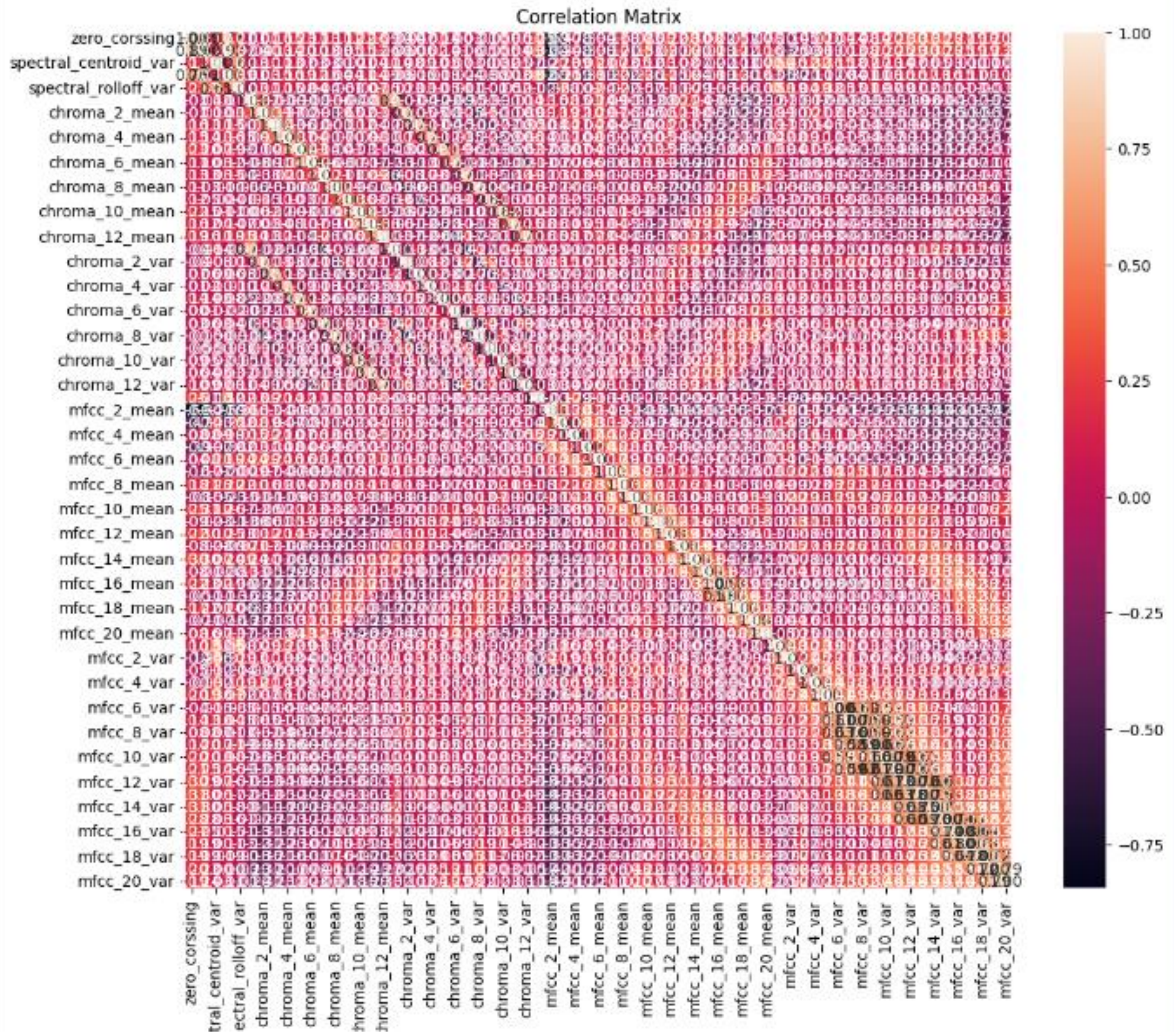
$$-5.108 = 0.158 b \Rightarrow b \simeq 3.2$$

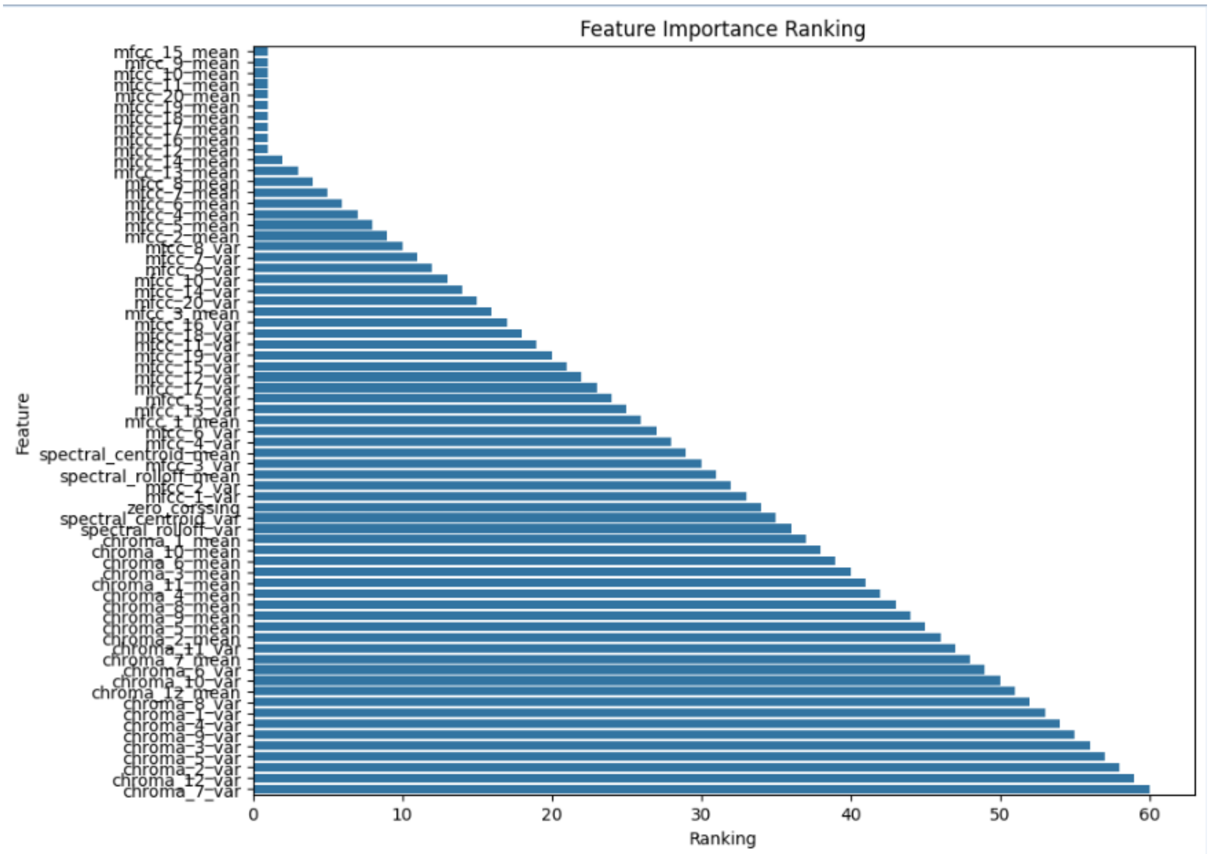
می‌توانیم تابع نگاشت $\varphi(x_1, x_2) = x_1^2 - x_2^2$ را معرفی کنیم. این نگاشت داده‌ها را از فضای دوبعدی به فضای یک بعدی منتقل کرده و دسته‌بندی خطی را امکان پذیر می‌کند.

پرسش دو

ستون های غیر عددی را از دیتاست حذف کردیم و کرولیشن ماتریکس را برای بقیه رسم کردیم.

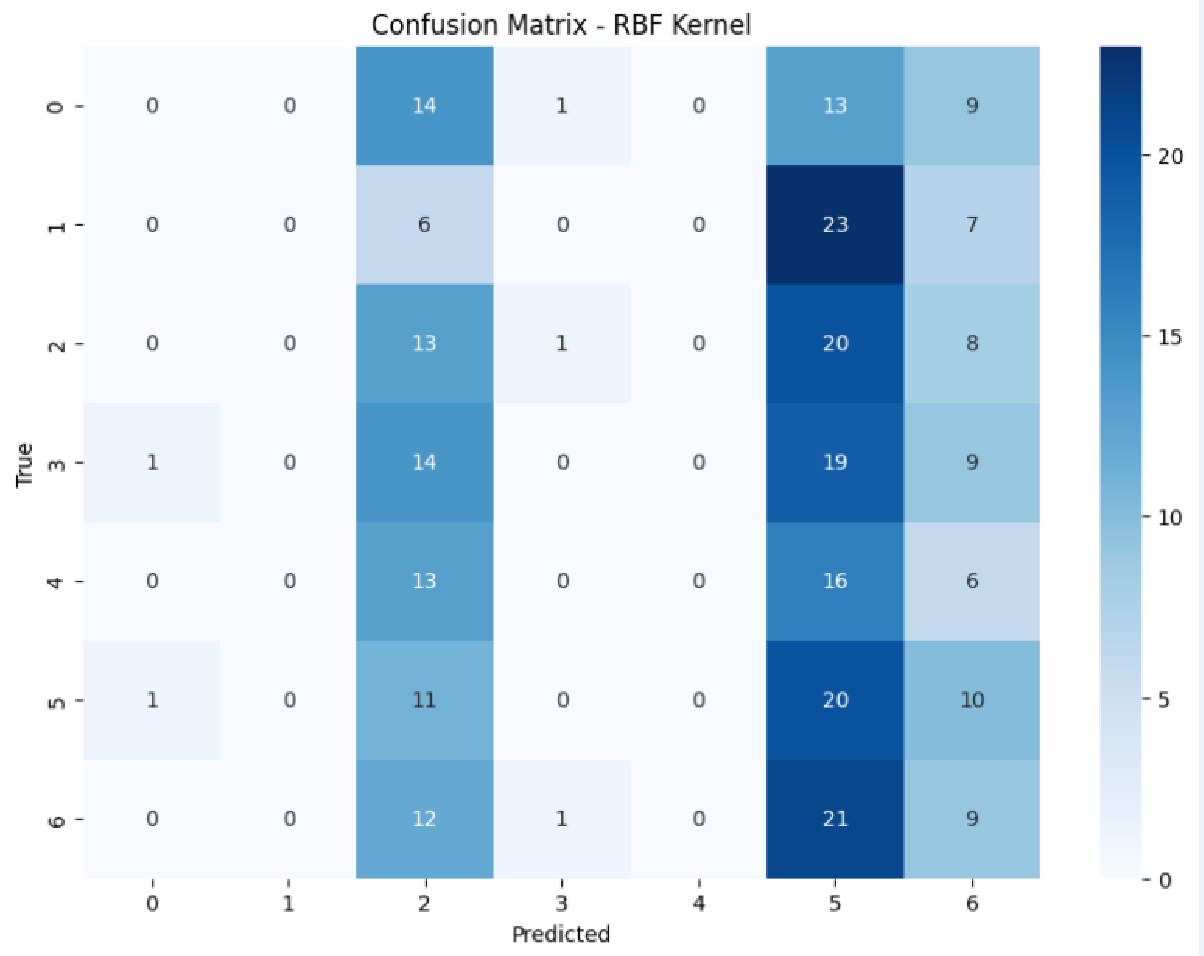

Class Distribution in Testing Data
Frequency
0
10
20
30
40
3
6
2
5
0
1
4
Class


Correlation Matrix
zero_corssing
spectral_centroid_var
spectral_rolloff_var
chroma_2_mean
chroma_4_mean
chroma_6_mean
chroma_8_mean
chroma_10_mean
chroma_12_mean
chroma_2_var
chroma_4_var
chroma_6_var
chroma_8_var
chroma_10_var
chroma_12_var
mfcc_2_mean
mfcc_4_mean
mfcc_6_mean
mfcc_8_mean
mfcc_10_mean
mfcc_12_mean
mfcc_14_mean
mfcc_16_mean
mfcc_18_mean
mfcc_20_mean
mfcc_2_var
mfcc_4_var
mfcc_6_var
mfcc_8_var
mfcc_10_var
mfcc_12_var
mfcc_14_var
mfcc_16_var
mfcc_18_var
mfcc_20_var
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75



سپس با هسته خطی و rbf مدل svm را آموزش می دهیم.

هسته rbf: در لینک دوم کولب

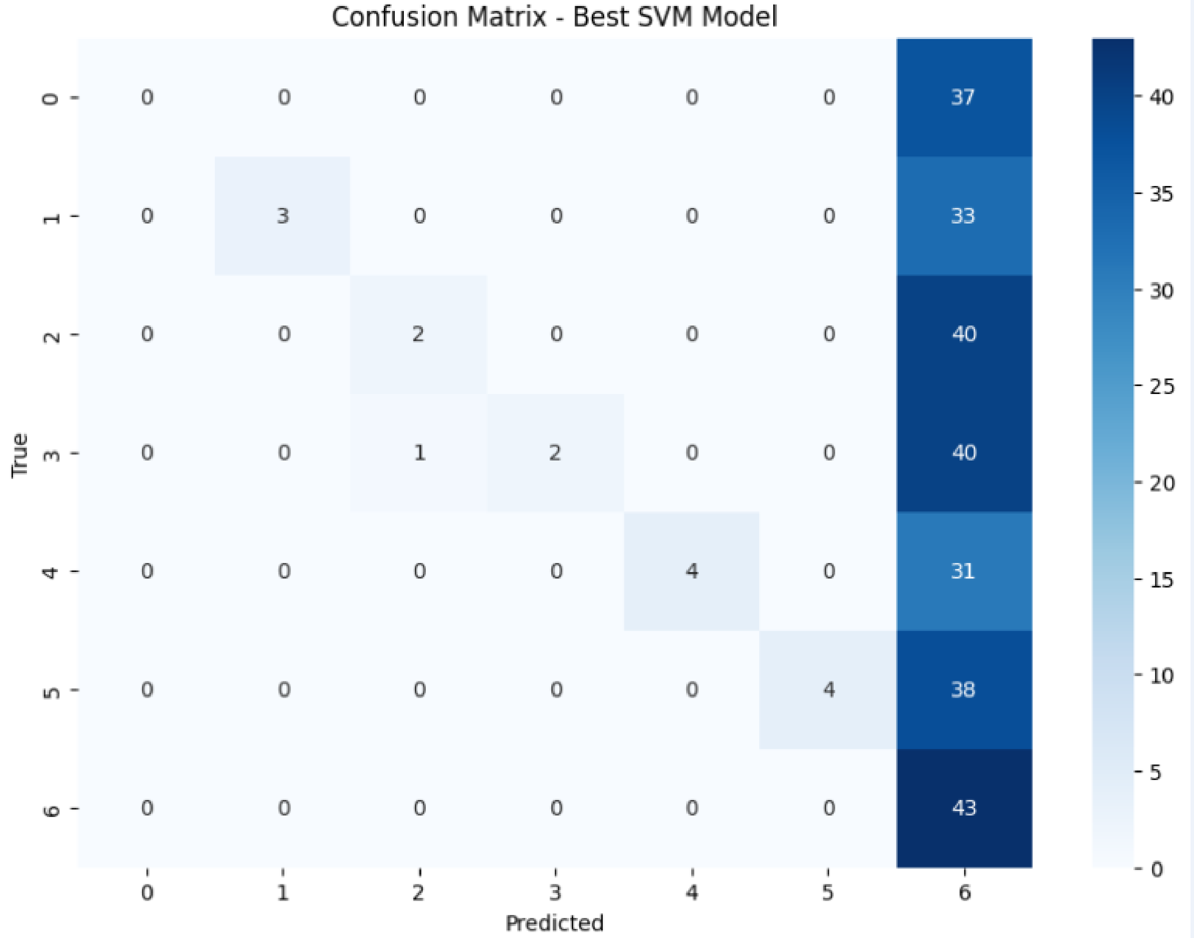



```
Classification Report - RBF Kernel
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        37
           1       0.00      0.00      0.00        36
           2       0.16      0.31      0.21        42
           3       0.00      0.00      0.00        43
           4       0.00      0.00      0.00        35
           5       0.15      0.48      0.23        42
           6       0.16      0.21      0.18        43

    accuracy                           0.15       278
   macro avg       0.07      0.14      0.09       278
weighted avg       0.07      0.15      0.09       278
```

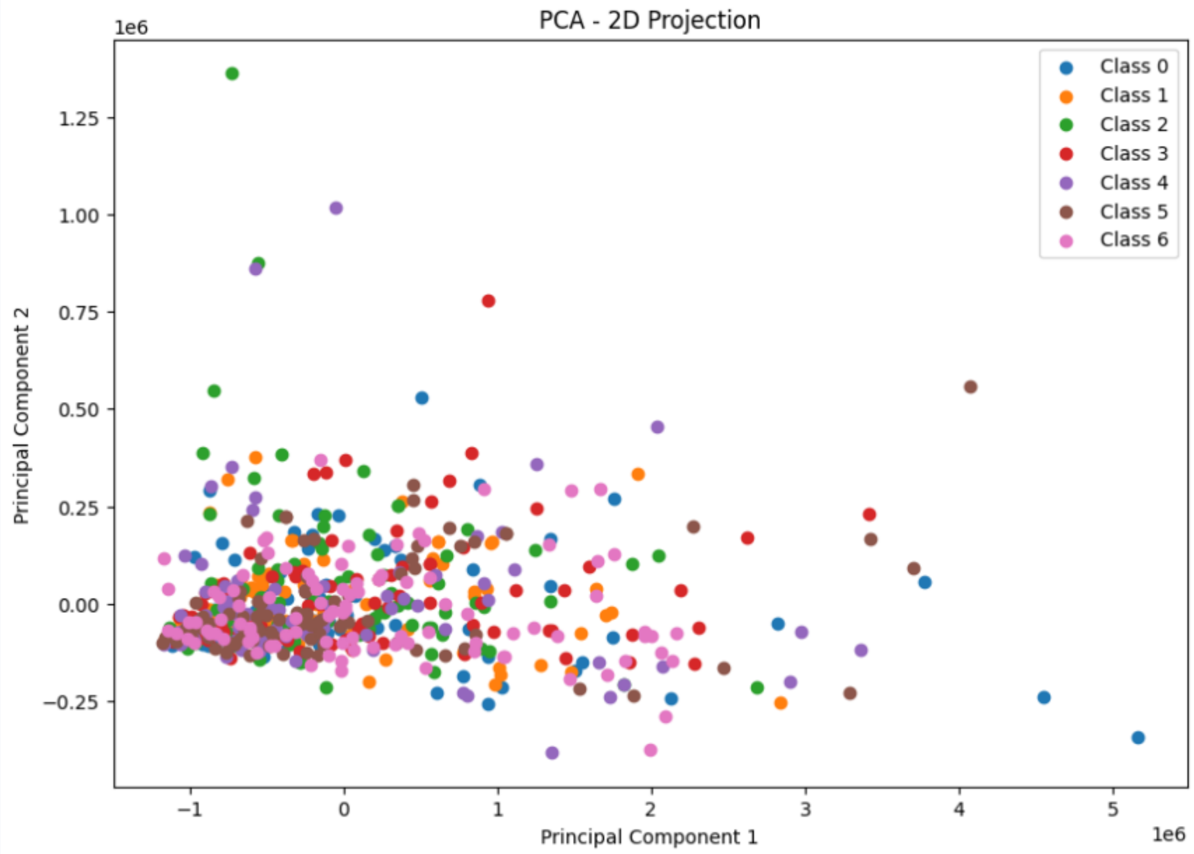
قسمت ج در لینک دوم کولب

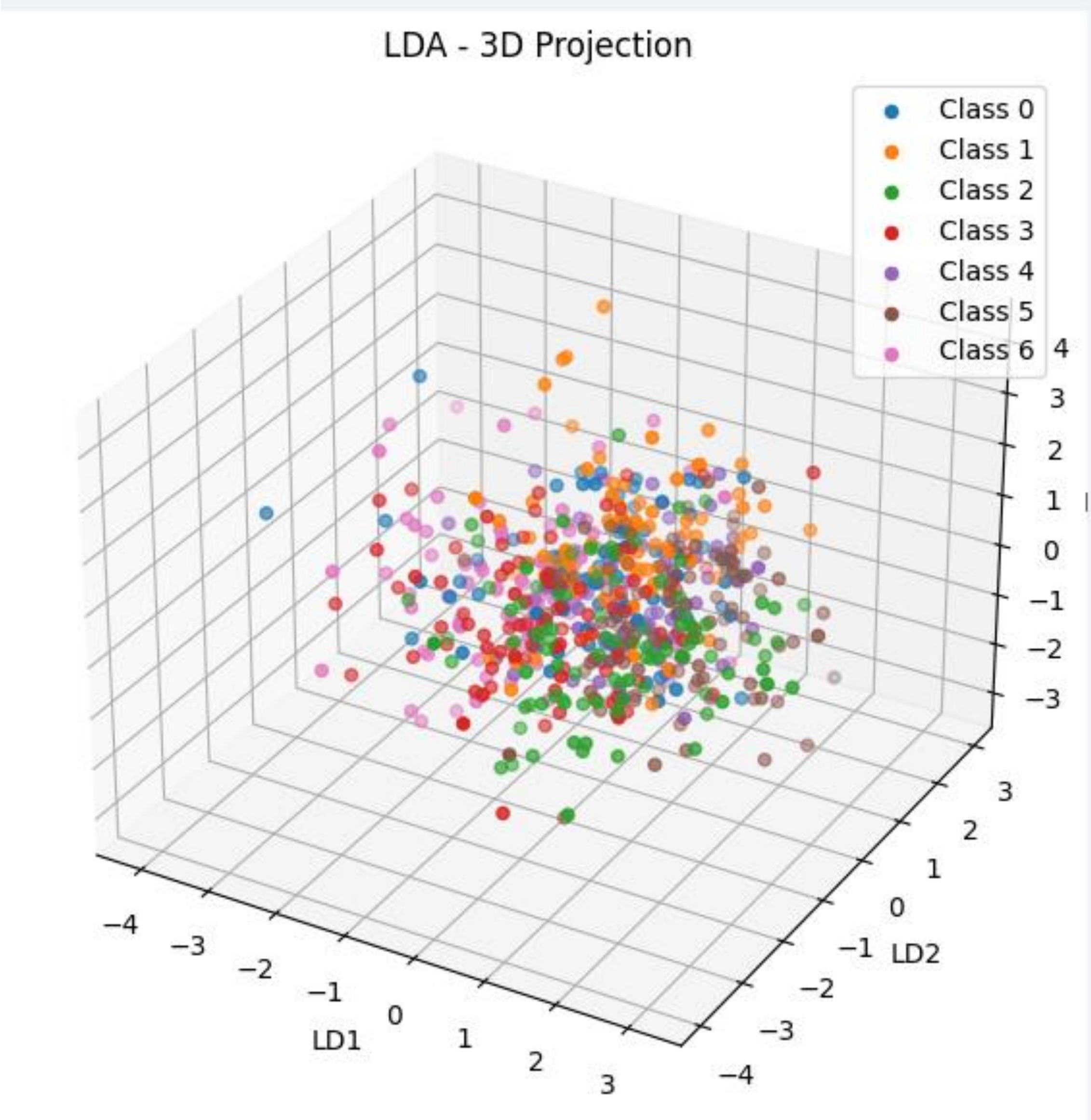


Classification Report - Best SVM Model

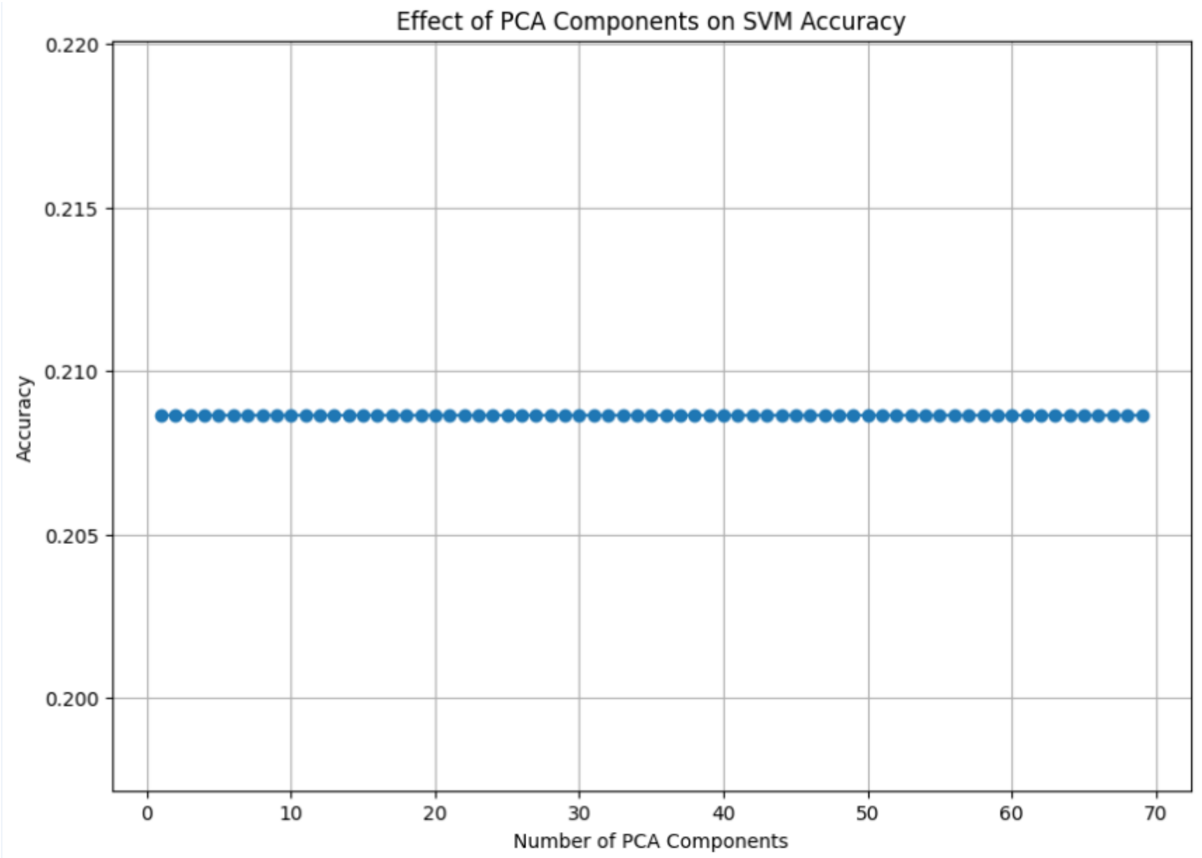
	precision	recall	f1-score	support
0	0.00	0.00	0.00	37
1	1.00	0.08	0.15	36
2	0.67	0.05	0.09	42
3	1.00	0.05	0.09	43
4	1.00	0.11	0.21	35
5	1.00	0.10	0.17	42
6	0.16	1.00	0.28	43
accuracy			0.21	278
macro avg	0.69	0.20	0.14	278
weighted avg	0.69	0.21	0.14	278

بخش د


PCA - 2D Projection
1e6
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
Principal Component 2
−1
0
1
2
3
4
5
1e6
Principal Component 1
Class 0
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6



LDA - 3D Projection
Class 0
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
LD1
LD2



در لینک سومی کولب به اسم final_ml2کل کد یکجا اورده شده فقط ران شدنش طول کشید عکساشو نذاشتم

پرسش سه

$(A_1, Right) \to A_2 \quad \hat{Q}(A_1, Right) = 0 + 0.9 \times max \hat{Q}(A_2, a') = 0$

$(A_2, Right) \to A_3 \quad \hat{Q}(A_2, Right) = 0 + 0.9 \times max \hat{Q}(A_3, a') = 0$

$(A_3, Right) \to B_3 \quad \hat{Q}(A_3, Right) = 0 + 0.9 \times max \hat{Q}(B_3, a') = 0$

$(B_3, Right) \to B_4 \quad \hat{Q}(B_3, Right) = -10 + 0.9 \times max \hat{Q}(B_4, a') = -10$

$(C_2, Left) \to C_1 \quad \hat{Q}(C_2, Left) = 0 + 0.9 \times max \hat{Q}(C_1, a') = 0$

$(C_1, \overset{R}{\cancel{up}}) \to B_1 \quad \hat{Q}(C_1, up) = 0 + 0.9 \times max \hat{Q}(B_1, a') = 0$

$(B_1, up) \to A_1 \quad \hat{Q}(B_1, up) = 0 + 0.9 \times max \hat{Q}(A_1, a') = 0$

$(A_1, a') = 0$

$(A_1, right) \to A_3 \quad \hat{Q}(A_1, right) = 0 + 0.9 \times max \hat{Q}(A_3, a') = 0$

$(A_3, right) \to A_4 \quad \hat{Q}(A_3, right) = 10 + 0.9 \times max \hat{Q}(A_4, a') = 10$

$(C_3, \underset{R}{\cancel{up}}) \to B_3 \quad \hat{Q}(C_3, up) = \cancel{5}\, 0 + 0.9 \times max \hat{Q}(B_3, a') = 0.9 \times -10 = -9$

$(B_3, \underset{R}{\cancel{up}}) \to A_3 \quad \hat{Q}(B_3, up) = -10 + 0.9 \times max \hat{Q}(A_3, a') = -10 + 0.9 \times 10 = -1$

$(A_3, right) \to A_4 \quad \hat{Q}(A_3, right) = +10 + 0.9 \times \underset{0}{max} \hat{Q}(A_4, a') = 10$

Q values $= \hat{Q}(A_1, R) = 0 \quad \hat{Q}(A_2, R) = 0 \quad \hat{Q}(A_3, R) = 10$

$\hat{Q}(B_3, R) = -1$, $\hat{Q}(B_4, R) = -10 \quad \hat{Q}(C_2, L) = 0$

$\hat{Q}(C_1, R) = 0 \quad \hat{Q}(B_1, U) = 0 \quad \hat{Q}(A_3, R) = 10$

$\hat{Q} = (A_4, R) = 10 \quad \hat{Q}(C_3, R) = -9$

برای حل این سؤال باید بهترین عملکرد را با بالاترین مقدار Q برای هر حالت تعیین کنیم

$A \rightarrow \rightarrow \rightarrow \; +10$

$B \rightarrow \; X \leftarrow \; -10$

$C \rightarrow \leftarrow \leftarrow$

این سیاست بهینه نیست چون ممکن است ربات در B_4 قرار بگیرد و پاداش منفی زیادی دریافت کند. یک سیاست بهتر از رفتن به B_4 تا حد امکان اجتناب کرده و به A_4 هدایت می‌شود.

⇐ سیاستی که از بالاترین مقدار Q برای هر حالت مشتق شده بهینه نیست. یک سیاست بهتر شامل اجتناب کامل از B_4 به دلیل پاداش منفی قابل توجه آن است.