

۱۳۰۷

دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده مهندسی برق - گرایش کنترل

پایان ترم

یادگیری ماشین

نگارش

فاطمه امیری

۴۰۲۰۲۴۲۴

لینک گوگل کولب

لینک گیت هاب

استاد مربوطه

جناب آقای دکتر علیاری

تیر ماه ۱۴۰۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فهرست مطالب

پرسش یک ۱

(آ) ۱

(ب) ۴

سوالات هماهنگ نشده تشریحی ۶

(ج) ۶

پرسش دو ۱۰

(آ) ۱۰

(ب) ۲۲

(ج) ۲۲

(د) ۲۳

(ه) ۲۹

پرسش سه ۳۲

پیش یک

در این بخش در ابتدا به بررسی دو سوال هماهنگ شده می پردازیم.

(T

این سوال که اثبات رابطه زیر را خواسته است، از [مقاله](#) [فرنسی استفاده](#) میشود.

فاطمه امیری

$$L(w, b, \lambda, R, \epsilon) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i (w^T x_i + b) - 1 + \epsilon_i) - \sum_{i=1}^N \gamma_i \epsilon_i + C \sum_{i=1}^N \epsilon_i^2$$

با قرار دادن این قید در لاگرانژین $\frac{\partial L}{\partial w}$ و $\frac{\partial L}{\partial b}$...

$0 \leq \alpha_i \leq C$ باید برقرار باشد. ثابت کنید اگر بخواهیم این لاگرانژین را \min کنیم، به شرط $0 \leq \alpha_i \leq C$ می‌رسیم که

در اینجا C میزان Soft margin بودن را متعوض می‌کند.

این مقاله پژوهشی با عنوان "شبکه‌های پشتیبانی بردار" توسط کورینا کورتس و ولادیمیر واپنیک در سال ۱۹۹۵ منتشر شده است. این مقاله مفهوم شبکه‌های پشتیبانی بردار را به عنوان یک ماشین یادگیری نوین برای مسائل دسته‌بندی دوگروهی معرفی می‌کند. ایده اصلی شامل نگاشت بردارهای ورودی به یک فضای ویژگی با بعد بالا و ساختن یک سطح تصمیم‌گیری خطی است که توانایی بالایی در تعمیم‌دهی دارد.

این روش برای داده‌های پیچیده و غیرقابل جداسازی نیز مقاوم است و در مقایسه با الگوریتم‌های یادگیری سنتی عملکرد بهتری نشان می‌دهد. جنبه‌های کلیدی مقاله شامل فرمول‌بندی ریاضی، استراتژی‌های بهینه‌سازی برای حاشیه بین کلاس‌ها، و مدیریت داده‌های غیرقابل جداسازی است. مقاله همچنین به ملاحظات محاسباتی و پیاده‌سازی عملی شبکه‌های پشتیبانی‌بردار می‌پردازد و پیشرفتی مهم در حوزه یادگیری ماشین را ارائه می‌دهد.

حال برای اثبات این سوال از این مقاله کمک گرفتیم و اثبات در صفحه بعد آمده است.

¹ <https://link.springer.com/article/10.1007/Bf00994018>

² Support-Vector Networks

فاطمه امیری

$$L(w, b, \alpha, r, \epsilon) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i (w^T x_i + b) - 1 + \epsilon_i) - \sum_{i=1}^N r_i \epsilon_i + c \sum_{i=1}^N \epsilon_i^2$$

... و $\frac{\partial L}{\partial w}$ و $\frac{\partial L}{\partial b}$ را لگرانژین

با قرار دادن این قید در لگرانژین باید برقرار باشد. ثابت کنید اگر بخواهیم این لگرانژین را \min کنیم، به شرط $0 \leq \alpha_i \leq c$ می‌رسیم که در اینجا c وزن Soft margin بود. margin را متخض می‌کند.

ثبات:

$$\Phi = \frac{1}{2} w \cdot w + c \left(\sum_{i=1}^N \epsilon_i \right)^2$$

$$\begin{cases} y_i (x_i \cdot w + b) \geq 1 - \epsilon_i & i=1, \dots, N \\ \epsilon_i \geq 0 & i=1, \dots, N \end{cases} \quad \text{شرط:}$$

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

که $r_i \geq 0$ و $\alpha_i \geq 0$ است ← ضرایب لگرانژ

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

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

$$\frac{\partial L}{\partial \epsilon_i} = 2c \sum_{i=1}^N \epsilon_i - \alpha_i - r_i = 0 \rightarrow \alpha_i + r_i = 2c \sum_{i=1}^N \epsilon_i = \delta$$

$$\Rightarrow \sum_{i=1}^N \epsilon_i = \frac{\delta}{2c}$$

بافرض $r_i \epsilon_i = 0$ و $\epsilon_i \geq 0$ ← داریم: $\alpha_i = 2c \sum_{i=1}^N \epsilon_i$

حال w را در L جای‌گذاری می‌کنیم و داریم:

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

حالا L را به صورت دیگری می‌نویسیم، چرا که $\sum_{i=1}^N \alpha_i y_i = 0$ و $w = \sum_{i=1}^N \alpha_i y_i x_i$ در فرم دی‌یابی مانده ظاهر شده؛ پس فرم دیگر L را در ضمیمه بعد می‌نویسیم:

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

باتوجه به مقاله فرانس، فرم دوگان این تابع هزینه به صورت زیر نوشته می شود:

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

باتوجه به اینکه $\alpha_i \geq 0$ ، $\Lambda = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}$ ، $D = \text{diag}(x_i \cdot x_j y_i y_j)$

داریم: $w(\Lambda, \delta) = \Lambda^T 1 - \left(\frac{1}{\gamma} \Lambda^T D \Lambda + \frac{\delta^2}{\gamma C} \left(1 - \frac{1}{\gamma} \right) \right)$ (I)

بشرط های

$$\Lambda^T \gamma = 0 , \Lambda + R = S1 , \Lambda \geq 0 , R \geq 0$$

باتوجه به اینکه $R \geq 0$ ، $\Lambda + R = S1$ می توانیم بنویسیم:

$$\Lambda = S1 - R \xrightarrow{\geq 0} \Lambda \leq S1$$

$\Lambda \geq 0$ پس داریم $0 \leq \Lambda \leq S1$

حال با فرض اینکه $\delta = \alpha_{\max} = \max(\alpha_1, \dots, \alpha_N)$ و با جای گذاری در معادله تابع هزینه (I) داریم:

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

باتوجه به نتایج پژوهش های گزارش شده در مقاله برای حل این مسئله QP دارای $K=2$ دینفر گرفته شده است و باتوجه به این، w به صورت زیر بدست می آید:

$$w(\Lambda) = \Lambda^T 1 - \frac{1}{\gamma} \Lambda^T D \Lambda$$

$$\text{شرط 2 (میرد)} \begin{cases} 0 \leq \Lambda \leq C1 \\ \Lambda^T \gamma = 0 \end{cases}$$

از قید اول داریم $0 \leq \Lambda \leq C1 \longrightarrow 0 \leq \alpha_i \leq C$

با این قید w باجهینه است و در نتیجه L نیز باجهینه است.

پس ثابت می شود قید $0 \leq \alpha_i \leq C$ تابع هزینه را باجهینه می کند.

(ب)

این سوال در واقع اثبات رابطه زیر را خواسته است :

فاصله آماری

برای اینکه بالانس نبوده کلاس ها را طرک کنند ، تأثیر فراوانی هر کلاس را کلاس در نظر باشد :

$$S_B = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu) \times N_i$$

↓ S_B : مطلوب
↑ S_B : برگشتی دانی

↓ S_W : مطلوب
↑ S_W : برگشتی دانی

باید به دنبال یک ماتریس باشیم max کنیم که سخت است پس با

اثبات کنید Scatter Total داده که طبق رابطه زیر بدست می آید :

$$\sum_{Total} \Rightarrow S_P = S_W + S_B \Rightarrow$$

↓ S_{Total} : $S_{Total} = S_B + S_W$

برای اثبات این سوال از کتاب ["Pattern Classification"](#) نوشته ریچارد او. دودا، پیترو ای. هارت، و دیوید جی. استورک کمک گرفته شده است که این کتاب یکی از مراجع اصلی و معتبر در زمینه طبقه بندی الگوها و یادگیری ماشین است. این کتاب که اولین بار در سال ۱۹۷۳ منتشر شد و در سال ۲۰۱۲ به روزرسانی شد، به بررسی مبانی تئوری و عملی تکنیک های طبقه بندی، از جمله روش های بیزین، شبکه های عصبی، ماشین های بردار پشتیبان (SVM) و خوشه بندی می پردازد. همچنین، مباحثی نظیر کاهش ابعاد و تحلیل مؤلفه های اصلی (PCA) را پوشش می دهد. این کتاب با ارائه مثال ها و تمرین های متعدد ما را به عمق مفاهیم کلیدی و کاربردی این حوزه هدایت می کند و به عنوان یک منبع جامع و قابل اعتماد مورد استفاده قرار می گیرد.

در صفحه بعد به اثبات رابطه خواسته شده در سوال می پردازیم.

فاطمه امیری

برای اینکه بالانس نبودن کلاس ها را حل کنند، تأثیر فراوانی هر کدام از کلاس در نظر باشد:

$$S_B = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu) \times N_i$$

Between Scatter

مطلوب: S_B و S_W
برگشتی میانی
برگشتی داخلی

باید به دنبال یک ماتریس باشیم \max کنیم که سخت است پس ما

$$\Rightarrow S_P = S_W^{-1} S_B \Rightarrow$$

نسبت کنید Scatter Total داده که طبق رابطه زیر بدست می آید:

$$S_{Total} = S_B + S_W$$

$$S_W = \sum_{i=1}^c S_i, \quad S_i = \sum_{j=1}^{N_i} (x_j - \mu_i)(x_j - \mu_i)^T$$

$$\Rightarrow \mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j$$

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

x : تمام داده ها، c : تعداد کلاس ها

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

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

$$\Rightarrow S_T = \sum_{i=1}^c (x - \mu)(x - \mu)^T = \sum_{i=1}^c \sum_{j=1}^{N_i} (x_j - \mu_j + \mu_i - \mu)(x_j - \mu_j + \mu_i - \mu)^T =$$

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

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

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

$$\textcircled{I} \Rightarrow \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T = S_W + S_B = S_T$$

پس رابطه ثابت شد.

سوالات هماهنگ نشده تشریحی

(ج)

پرسش یک (ج)

$$x_1 = \{ \overbrace{(1, 1)}^{\text{کلاس}}, 1 \} \rightarrow$$

$$\varphi(x) = (\overbrace{x_1 + x_2}^{\text{کلاس}}, \overbrace{x_1 - x_2}^{\text{کلاس}})$$

مجاوبت پیدا می کنند به فضای تصویر
(mapping طبق رابطه $\varphi(x)$)

$$\varphi(x_1) = (2, 0)$$

$$\varphi(x_2) = (5, 1)$$

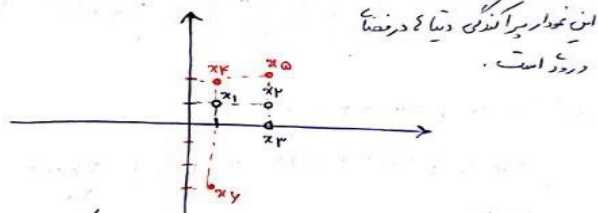
$$\varphi(x_3) = (4, 2)$$

$$\varphi(x_4) = (5, -1)$$

$$\varphi(x_5) = (8, 0)$$

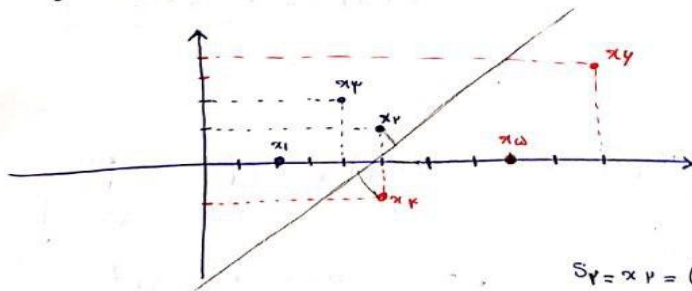
$$\varphi(x_6) = (10, 4)$$

$$\left. \begin{array}{l} x_1 \rightarrow 1 \\ x_2 \rightarrow 1 \\ x_3 \rightarrow 1 \\ x_4 \rightarrow -1 \\ x_5 \rightarrow -1 \\ x_6 \rightarrow -1 \end{array} \right\} \begin{array}{l} \text{کلاس مثبت} \\ \text{مقرض مثبت} \end{array}$$



این نمودار برای اندکی دیتا در فضای درود است.

نمونه داده می کنیم که من توهم همین جور یک خط می کشیم و این را جدا کنیم پس از $\varphi(x)$ استفاده می کنیم که یک تبدیل غیر خطی است و داده را به فضای من می برد که من توهم تشکیل کنیم کلاس را.



(۱)
برای جدا کردن این دو دسته به صورت زیر هستند:

$$S_2 = x_2 = (5, 1) \rightarrow \text{کلاس } +1$$

$$S_1 = x_4 = (5, -1) \rightarrow \text{کلاس } -1$$

چون این نقاط به صورت خطی به هم نمی افتند.

ب) پیدا کردن ستادیرجینه وزن و بایاس

$$\begin{cases} \alpha_1 \phi_1(s_1) \cdot \phi(s_1) + \alpha_2 \phi_1(s_2) \cdot \phi(s_1) = -1 \\ \alpha_1 \phi_1(s_1) \cdot \phi(s_1) + \alpha_2 \phi_1(s_2) \cdot \phi(s_2) = +1 \end{cases}$$

$$\Rightarrow \begin{cases} \alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_1 = -1 \\ \alpha_1 \tilde{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 = +1 \end{cases}$$

حالا باید α ها که وزن هستند را حساب کنیم

لازم به ذکر است که مقدار

بایاس را 1 در نظر میگیریم.

$$\begin{cases} \tilde{s}_1 \cdot \tilde{s}_1 \rightarrow \begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix} = 25 + 1 + 1 = 27 \\ \tilde{s}_2 \cdot \tilde{s}_1 \rightarrow \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix} = 25 + (-1) + 1 = 25 \\ \tilde{s}_2 \cdot \tilde{s}_2 \rightarrow \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix} = 25 + 1 + 1 = 27 \end{cases}$$

$$\Rightarrow \begin{cases} 27 \alpha_1 + 25 \alpha_2 = -1 \\ 25 \alpha_1 + 27 \alpha_2 = +1 \end{cases} \rightarrow \alpha_1 = \frac{1 - 27 \alpha_2}{25}$$

α ها ستادیرجینه وزن هستند

$$\frac{27 - 729 \alpha_2 + 625 \alpha_2}{25} = -1 \rightarrow 27 - 104 \alpha_2 = -25$$

$$-104 \alpha_2 = -52 \rightarrow \alpha_2 = 0.5$$

$$\alpha_1 = -0.5$$

$$\xrightarrow{\text{در آخر}} \tilde{w} = \sum_i \alpha_i \tilde{s}_i \rightarrow = -0.5 \begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix} + 0.5 \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$y = w x + b$$

$$b = 0, w = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} x \Rightarrow \text{hyperplane}$$

ج) از کرنل K استفاده می‌کنیم: مثل کرنل زیر:

$$(a^2 + 2ab + 1)$$

$$k(v, u) = (1 + uv)^2 = 1 + 2uv + v^2 u^2 = \underbrace{(1 \sqrt{2}u \quad u^2)}_{\phi(u)} \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}v} \\ v^2 \end{pmatrix}}_{\phi(v)}$$

حالا یک بار هم ستارها برداریم و بنویسیم در نظر می‌گیریم و داریم:

$$\begin{cases} \tilde{s}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} & y_1 = +1 \\ \tilde{s}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} & y_2 = -1 \\ \tilde{s}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} & y_3 = -1 \end{cases}$$

$$\alpha_1 \phi(s_1) \cdot \phi(s_2) + \alpha_2 \phi(s_2) \cdot \phi(s_1) + \alpha_3 \phi(s_3) \phi(s_1) = 1$$

$$\alpha_1 \phi(s_1) \phi(s_2) + \alpha_2 \phi(s_2) \phi(s_2) + \alpha_3 \phi(s_2) \phi(s_2) = -1$$

$$\alpha_1 \phi(s_1) \cdot \phi(s_3) + \alpha_2 \phi(s_2) \phi(s_3) + \alpha_3 \phi(s_3) \phi(s_3) = -1$$

$$w = \sum_i \alpha_i \tilde{s}_i \xrightarrow{\text{میزنویس}} \begin{cases} \alpha_1 = -1, 47 \\ \alpha_2 = 1, 25 \\ \alpha_3 = 1, 11 \end{cases}$$

$$w = \begin{pmatrix} 1, 4 \\ -1, 45 \\ -1, 1 \end{pmatrix} \rightarrow \text{میزنویس}$$

$$\Rightarrow w^T x + b = 0 \rightarrow (1, 4 \quad -1, 45) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (-1, 1) = 0$$

②

$$\varphi = (\pi_1^2 + \pi_2^2) - (\pi_1 - \pi_2) + \omega$$

$$\varphi_1(\pi_1) = 2$$

$$\varphi_2(\pi) = 2$$

$$\varphi_2(\pi) = 1$$

$$\varphi_3(\pi) = -1$$

$$\varphi_\omega(\pi) = -2$$

$$\varphi_4(\pi) = -1$$

رد شده و مرز دارند

پیش‌دو

آ. داده‌ها را به حداقل دو دسته آموزش و آزمون تقسیم کنید. توزیع کلاسی داده‌های آموزش و تست خود را از طریق نمودار میله‌ای نمایش دهید. ماتریس هم‌بستگی را رسم کنید و مهم‌ترین ویژگی‌ها را با استفاده از تحلیل یا استفاده از یک الگوریتم انتخاب ویژگی تعیین کنید و نمایش دهید.

ب. با استفاده از طبقه‌بند مبتنی بر SVM دستگاه‌های موسیقی را طبقه‌بندی کنید و نتایج را بصورت دقیق و کامل برای هر کلاس در قالب ماتریس درهم‌ریختگی (به صورت درصدی) نمایش دهید. تفاوت استفاده از دو دسته خطی و غیرخطی در نتایج را بررسی و تحلیل کنید. برای بخش‌های بعدی که نیاز به تکرار پیاده‌سازی این بخش پیدا می‌کنید، پیاده‌سازی به‌ازای یک دسته کافی است و نیازی به پیاده‌سازی و مقایسه روی دسته‌های مختلف نیست.

ج. با استفاده از روش‌هایی مانند GridSearch، حداقل یک فرایارامتر را انتخابی خود در بخش (ب) را تا جای ممکن بهینه و تنظیم کنید. همچنین، لازم است که در یافتن مقدار بهینه فرایارامتر انتخابی خود از روش K-Fold Cross Validation بهره بگیرید. دقت داشته باشید که تنظیم فرایارامتر در کل این سوال فقط یک بار لازم است و می‌توانید در مراحل بعدی از مقدار (مقادیر) به‌دست‌آمده در همین بخش استفاده کنید.

د. از هر دو روش PCA و LDA برای کاهش بُعد استفاده کنید و نتیجه بصری را برای PCA در فضای دو بُعدی و برای LDA در فضای سه بُعدی نمایش دهید (هر کلاس با رنگ متفاوت). سپس، روی یک نمودار تاثیر انتخاب تعداد مولفه‌های PCA بر دقت الگوریتم طبقه‌بندی بخش (ب) را نمایش دهید.

ه. حداقل یک ویژگی جدید از داده‌ها استخراج کنید. با توجه به ماهیت صوتی منشأ داده‌ها چه جنس ویژگی‌ای استخراج کرده‌اید؟ در ادامه، یک شبکه اتوانکودر طراحی کنید و از آن به عنوان استخراج‌کننده ویژگی استفاده کنید و بخش (ب) را تکرار کنید و نتایج را گزارش و تحلیل کنید. نمودار اتلاف بخش آموزش و اعتبارسنجی مربوط به اتوانکودر خود را هم رسم کنید. از کدام قسمت(های) شبکه برای انجام فرآیند خواسته شده در این سوال استفاده کردید؟ چرا؟

و. فرض کنید آهنگ «به سکوت سرد زمان» استاد شجریان در اختیار شما قرار داده شده و آن را با استفاده از الگوریتم‌های ارزیابی آماده‌شده به داده‌های عددی موجود در این پیوند تبدیل کرده‌اید. آیا می‌توانید امکان این را ایجاد کنید که با یکی از مدل‌هایی که در بخش‌های قبلی آموزش داده‌اید، دستگاه موسیقی مربوط به این داده را تعیین کنید؟ پیاده‌سازی کرده و نتیجه را گزارش کنید.

ز. به نظرتان کاربرد این پروژه چه می‌تواند باشد و چه نوع مدل‌هایی در این کاربرد بهتر جواب می‌دهند؟ به عنوان نمره امتیازی می‌توانید از روشی دیگر برای طبقه‌بندی استفاده کنید و نشان دهید که در این روش جدید نتیجه بهتر شده است.

(۲)

گام‌های زیر را طی می‌کنیم :

بارگذاری و پیش‌پردازش داده‌ها

در ابتدا، داده‌های مربوط به مجموعه موسیقی از فایل MJMusicDataset.csv بارگذاری شدند. ستون‌های نام و ساز به دلیل عدم ضرورت برای تحلیل فعلی، حذف گردیدند. متغیر هدف dastgah به مقادیر عددی با استفاده از یک دیکشنری جایگزین شد که به ما امکان می‌دهد از این متغیر به عنوان هدف مدل‌های یادگیری ماشین استفاده کنیم.

تقسیم داده‌ها به دسته‌های آموزشی و آزمایشی

داده‌ها به نسبت ۸۰ به ۲۰ به دو دسته آموزشی و آزمایشی تقسیم شدند. این تقسیم‌بندی به ما کمک می‌کند تا مدل‌هایمان را بر روی داده‌های آموزشی آموزش دهیم و سپس عملکرد آن‌ها را بر روی داده‌های آزمایشی ارزیابی کنیم.

تحلیل توزیع داده‌های آموزشی

برای بررسی توزیع داده‌های آموزشی، از نمودار میله‌ای استفاده شد. این نمودار نشان می‌دهد که تعداد نمونه‌های هر کلاس در داده‌های آموزشی چگونه توزیع شده است، که به ما کمک می‌کند از تعادل یا عدم تعادل داده‌ها آگاه شویم.

حلیل همبستگی ویژگی‌ها

برای بررسی روابط بین ویژگی‌های مختلف داده‌ها، ماتریس همبستگی نمایش داده شد. این ماتریس نشان می‌دهد که چگونه ویژگی‌های مختلف با هم مرتبط هستند. همبستگی بالا بین ویژگی‌ها ممکن است نشان‌دهنده اطلاعات تکراری باشد که در مرحله انتخاب ویژگی‌ها باید مورد توجه قرار گیرد.

کاهش ابعاد با LDA و t-SNE

برای کاهش ابعاد داده‌ها و تجسم بهتر آن‌ها، از تحلیل تشخیص خطی (LDA) و t-SNE استفاده شد. LDA برای کاهش ابعاد به تعداد محدودی از مؤلفه‌های خطی که بیشترین تمایز بین کلاس‌ها را ایجاد می‌کنند، استفاده شد. سپس t-SNE برای کاهش بیشتر ابعاد و تجسم داده‌ها در فضای دو بعدی به کار رفت.

استانداردسازی داده‌ها

برای بهبود عملکرد مدل‌های یادگیری ماشین، ویژگی‌ها با استفاده از استانداردسازی استانداردسازی شدند. استانداردسازی به ما کمک می‌کند تا تمامی ویژگی‌ها دارای مقیاس مشابهی باشند و هیچ یک تأثیر بیشتری بر مدل‌ها نداشته باشند.

آموزش مدل SVM

دو مدل SVM با کرنل‌های خطی و غیرخطی آموزش داده شدند. مدل با کرنل خطی یک SVM خطی بود که برای دسته‌بندی داده‌ها استفاده شد. مدل با کرنل غیرخطی، یک SVM با کرنل RBF بود که به دلیل انعطاف‌پذیری بیشتر در جداسازی کلاس‌ها، استفاده گردید. ماتریس درهم‌ریختگی و گزارش دسته‌بندی برای هر دو مدل محاسبه و نمایش داده شدند. این ارزیابی‌ها نشان داد که مدل با کرنل غیرخطی عملکرد بهتری نسبت به مدل خطی دارد.

بهینه‌سازی مدل با GridSearchCV

برای بهبود عملکرد مدل SVM با کرنل خطی، از GridSearchCV استفاده شد. این روش به ما اجازه داد تا با جستجو در فضای پارامترهای مختلف، بهترین ترکیب پارامترها را پیدا کنیم. پارامترهای مورد جستجو شامل C و tol بودند. بهترین پارامترها و امتیاز اعتبارسنجی متقاطع به دست آمد و مدل نهایی با استفاده از این پارامترها آموزش داده شد.

جمع‌بندی

این گزارش شامل مراحل مختلفی از جمله بارگذاری داده‌ها، پیش‌پردازش، تحلیل توزیع و همبستگی داده‌ها، کاهش ابعاد، استانداردسازی، آموزش مدل‌های SVM با کرنل‌های مختلف، و بهینه‌سازی مدل بود. مدل‌های مختلف SVM با استفاده از پارامترهای بهینه‌شده آموزش داده شدند و نتایج ارزیابی شامل ماتریس درهم‌ریختگی و گزارش دسته‌بندی نمایش داده شدند.

مدل SVM با کرنل RBF عملکرد بهتری نسبت به مدل خطی داشت و بهترین پارامترها برای آن $C=0.6$ بود. بهینه‌سازی مدل با GridSearchCV نیز منجر به بهبود عملکرد مدل SVM با کرنل خطی شد. این تحلیل‌ها و نتایج به ما کمک کردند تا بتوانیم بهترین مدل را برای دسته‌بندی داده‌های موسیقی انتخاب کنیم و عملکرد آن را بهبود بخشیم.

پابتدا دیتا ست را لود میکنم :

	name	dastgah	instrument	zero_crossing	spectral_centroid_mean	spectral_centroid_var	spectral_rolloff_mean	spectral_rolloff_var	chroma_1_mean	chroma_2_mean	...	mfcc_11_var	#
0	ney-mahoor-ebrahimi.mp3	D_2	I_4	65545	1938.040517	303472.4474	2815.166310	1.754412e+06	0.515210	0.158007	...	482.619965	
1	Mohammad_Shojaei_nei_Mahoor.mp3	D_2	I_4	59788	1956.981873	774951.2677	2971.032035	2.082504e+06	0.209722	0.335278	...	435.338196	
2	Arash_Samimi_nei_Mahoor.mp3	D_2	I_4	85072	2735.525193	575671.7929	4268.615855	2.663909e+06	0.105867	0.108526	...	299.526794	
3	19 sarebaang mahoor.mp3	D_2	I_4	89900	2629.389833	406198.9319	4071.233715	1.005496e+06	0.153988	0.098157	...	305.978638	
4	Amjadian.mp3	D_2	I_4	58134	1659.262559	332341.1003	2789.041468	8.496746e+05	0.077795	0.036009	...	81.764854	
...	
921	04- Dastgahe Segah (Dariush Safvat).mp3	D_1	I_3	58713	1962.963101	271920.3527	3971.409319	1.610546e+06	0.333098	0.462040	...	81.730835	
922	setar.mp3	D_1	I_3	76880	2250.524645	239123.6804	3555.770165	2.224720e+06	0.273377	0.446400	...	130.240204	
923	بنداد بابایی - پیش از آغوش سه گانه	D_1	I_3	51127	1773.256221	210447.3008	3243.288722	2.186991e+06	0.432290	0.573922	...	83.879944	
924	index.html?VALID=TRUE&affid=43702&VALID=TRUE&q...	D_1	I_3	66481	1548.315758	519856.3787	2582.517717	1.773740e+06	0.108252	0.187435	...	416.817444	
925	index.html?VALID=TRUE&affid=20705&VALID=TRUE&q...	D_1	I_3	95739	2149.124135	144537.8972	3839.285118	6.873622e+05	0.172524	0.235960	...	127.713089	

926 rows x 72 columns

```
[CV] END .....C=1, gamma=1; total time=
0.1s
[CV] END .....C=1, gamma=1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.1; total time=
0.1s
[CV] END .....C=1, gamma=0.01; total time=
0.1s
[CV] END .....C=1, gamma=0.01; total time=
0.1s
[CV] END .....C=1, gamma=0.01; total time=
0.1s
[CV] END .....C=1, gamma=0.01; total time=
0.1s
[CV] END .....C=1, gamma=0.01; total time=
0.1s
```



```

[CV] END .....C=1, gamma=0.001; total time=
0.0s
[CV] END .....C=1, gamma=0.001; total time=
0.1s
[CV] END .....C=1, gamma=0.001; total time=
0.0s
[CV] END .....C=1, gamma=0.001; total time=
0.1s
[CV] END .....C=1, gamma=0.001; total time=
0.0s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.1; total time=
0.1s
[CV] END .....C=10, gamma=0.01; total time=
0.1s
[CV] END .....C=10, gamma=0.01; total time=
0.1s
[CV] END .....C=10, gamma=0.01; total time=
0.1s
[CV] END .....C=10, gamma=0.01; total time=
0.1s
[CV] END .....C=10, gamma=0.01; total time=
0.1s
[CV] END .....C=10, gamma=0.001; total time=
0.1s
[CV] END .....C=10, gamma=0.001; total time=
0.1s
[CV] END .....C=10, gamma=0.001; total time=
0.0s
[CV] END .....C=10, gamma=0.001; total time=
0.1s
[CV] END .....C=10, gamma=0.001; total time=
0.0s
[CV] END .....C=100, gamma=1; total time=
0.1s
[CV] END .....C=100, gamma=1; total time=
0.1s

```

```

[CV] END .....C=100, gamma=1; total time=
0.1s
[CV] END .....C=100, gamma=1; total time=
0.1s
[CV] END .....C=100, gamma=1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.1; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.01; total time=
0.1s
[CV] END .....C=100, gamma=0.001; total time=
0.1s
[CV] END .....C=100, gamma=0.001; total time=
0.1s
[CV] END .....C=100, gamma=0.001; total time=
0.1s
[CV] END .....C=100, gamma=0.001; total time=
0.1s
[CV] END .....C=100, gamma=0.001; total time=
0.1s
Best parameters: {'C': 100, 'gamma': 0.01}
Best cross-validation score: 0.4108108108108109

```

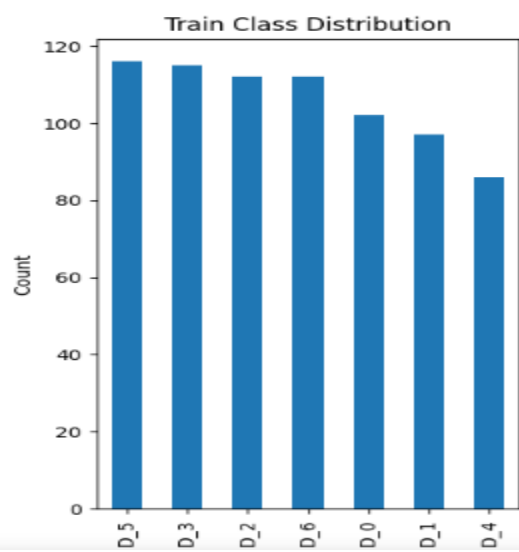
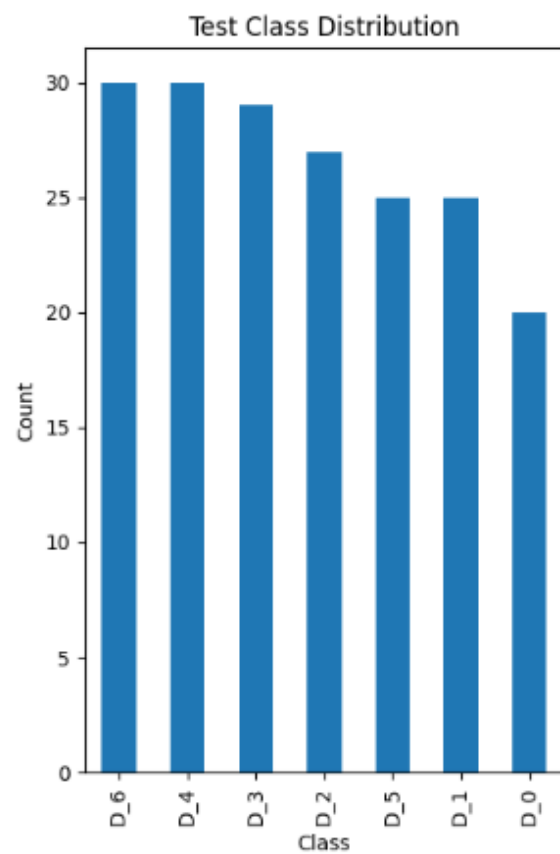
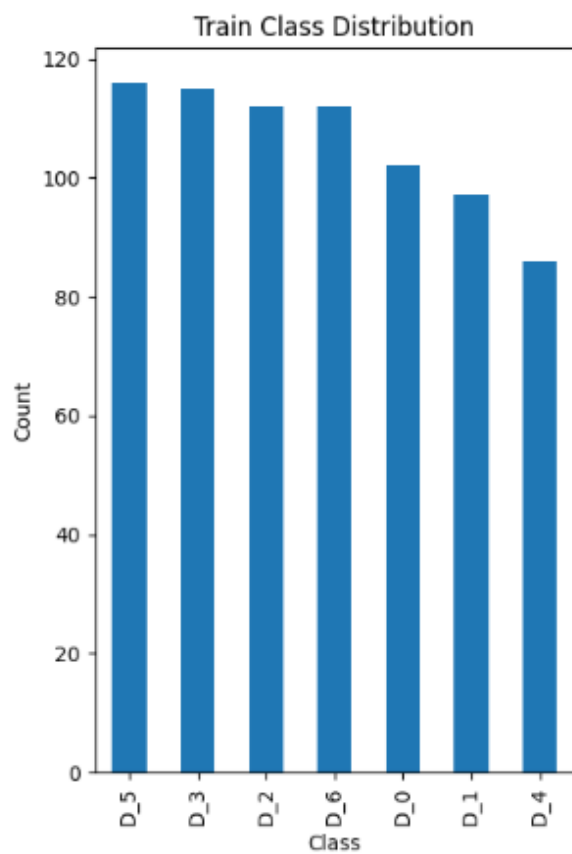
استخراج اطلاعات :

```

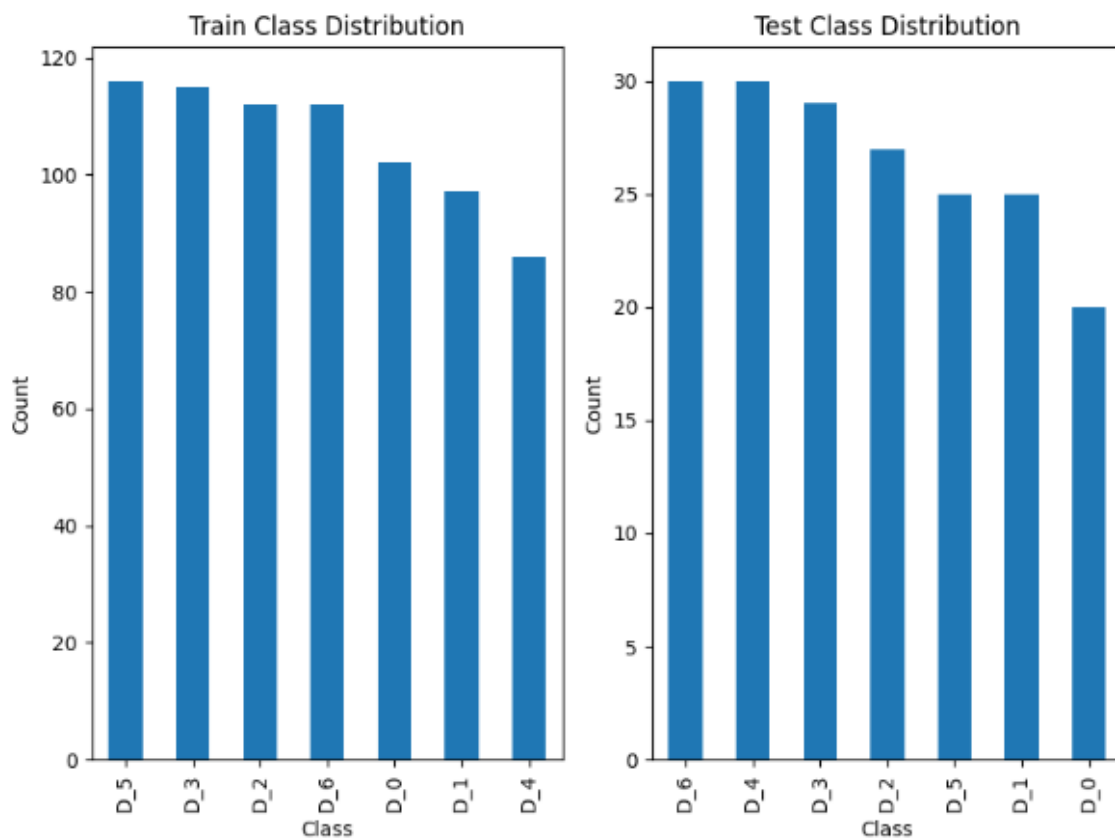
✓ [7] df.info()
06
↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 926 entries, 0 to 925
Data columns (total 72 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   name                                  926 non-null    object
1   dastgah                               926 non-null    object
2   instrument                             926 non-null    object
3   zero_crossing                         926 non-null    int64
4   spectral_centroid_mean                926 non-null    float64
5   spectral_centroid_var                 926 non-null    float64
6   spectral_rolloff_mean                 926 non-null    float64
7   spectral_rolloff_var                  926 non-null    float64
8   chroma_1_mean                         926 non-null    float64
9   chroma_2_mean                         926 non-null    float64
10  chroma_3_mean                         926 non-null    float64
11  chroma_4_mean                         926 non-null    float64
12  chroma_5_mean                         926 non-null    float64
13  chroma_6_mean                         926 non-null    float64
14  chroma_7_mean                         926 non-null    float64
15  chroma_8_mean                         926 non-null    float64
16  chroma_9_mean                         926 non-null    float64
17  chroma_10_mean                        926 non-null    float64
18  chroma_11_mean                        926 non-null    float64
19  chroma_12_mean                        926 non-null    float64
20  chroma_13_mean                        926 non-null    float64

```

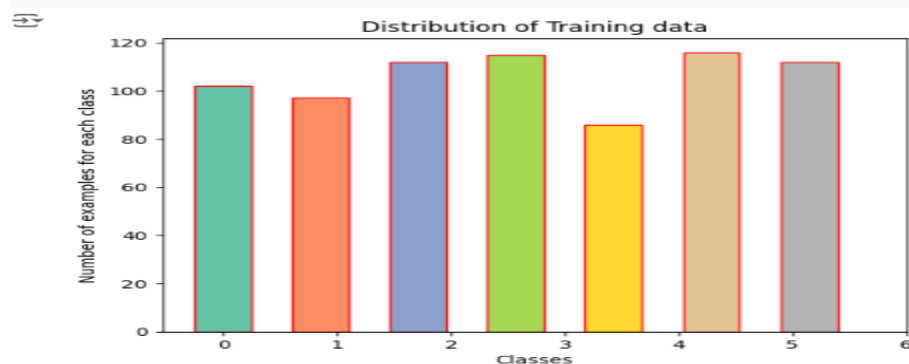
مشاهده نمودار های میله ای :

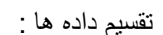


پراکندگی داده ها :

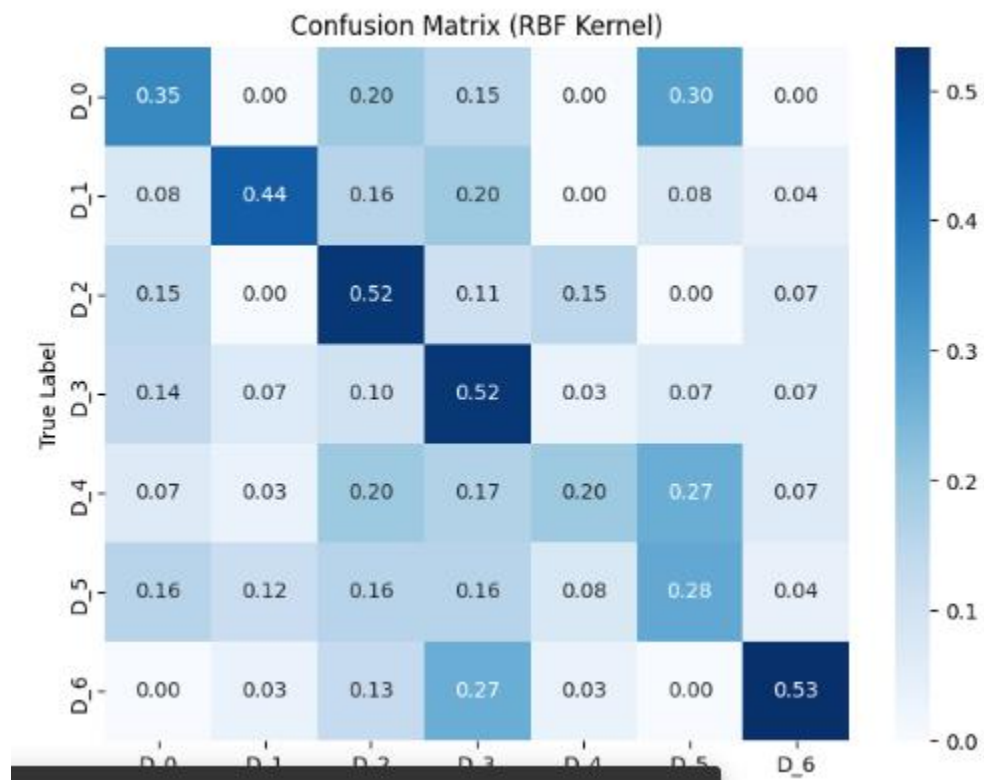
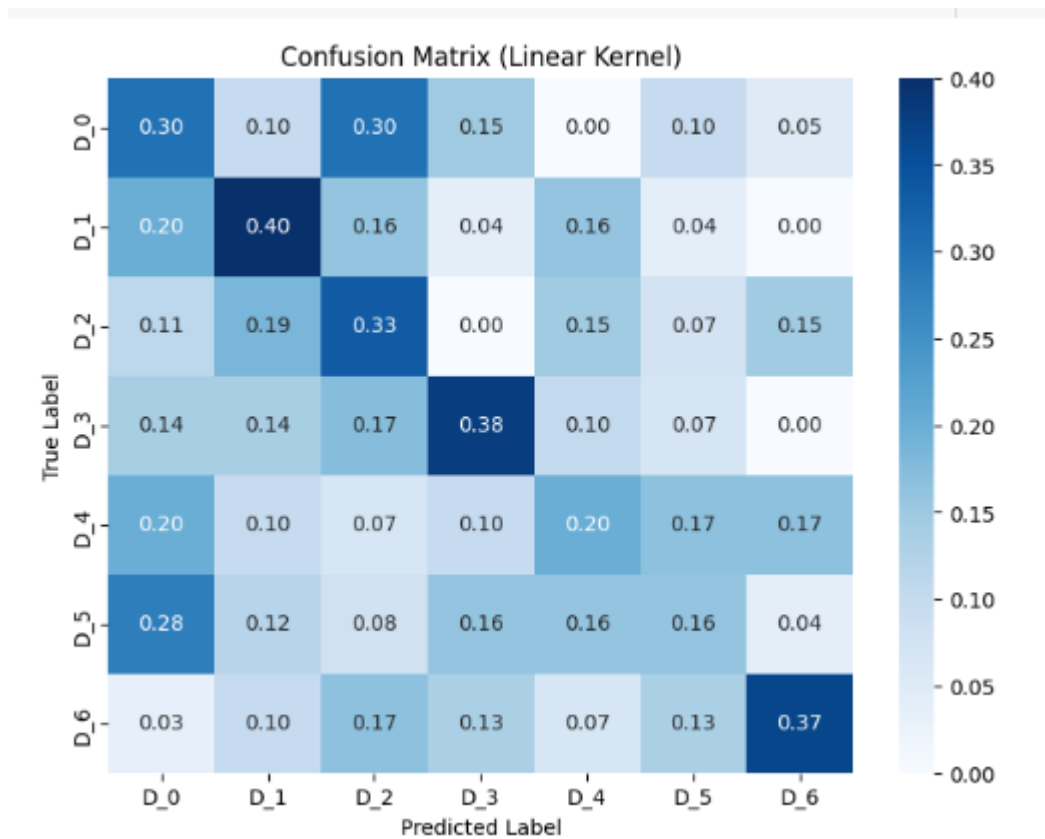


توزیع داده های بخش Train را میتوانیم مشاهده کنیم.





19



```
[13] Predicted Label
      Classification Report (Linear Kernel):
      precision    recall  f1-score   support

    D_0      0.19      0.30      0.23      20
    D_1      0.33      0.40      0.36      25
    D_2      0.27      0.33      0.30      27
    D_3      0.42      0.38      0.40      29
    D_4      0.26      0.20      0.23      30
    D_5      0.20      0.16      0.18      25
    D_6      0.50      0.37      0.42      30

 accuracy      0.31      186
 macro avg      0.31      0.31      0.30      186
 weighted avg    0.32      0.31      0.31      186

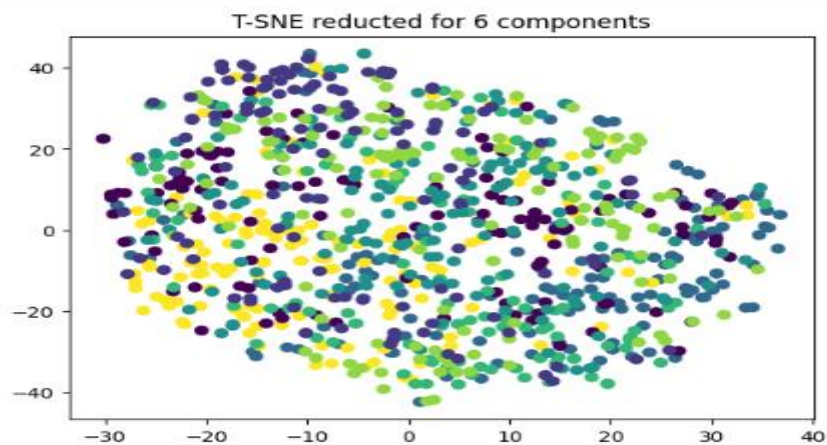
      Classification Report (RBF Kernel):
      precision    recall  f1-score   support

    D_0      0.30      0.35      0.33      20
    D_1      0.61      0.44      0.51      25
    D_2      0.36      0.52      0.42      27
    D_3      0.35      0.52      0.42      29
    D_4      0.43      0.20      0.27      30
    D_5      0.28      0.28      0.28      25
    D_6      0.67      0.53      0.59      30

 accuracy      0.41      186
 macro avg      0.43      0.41      0.40      186
 weighted avg    0.44      0.41      0.41      186
```

کاهش بعد :

```
plt.show()
```



```
from sklearn.preprocessing import StandardScaler
```


(ب)

دقت (Precision) و بازخوانی (Recall) به طور کلی، مدل با هسته RBF نسبت به مدل با هسته خطی عملکرد بهتری دارد. دقت و بازخوانی برای اکثر کلاس‌ها در مدل RBF بالاتر است، به ویژه برای کلاس‌های D_3 و D_6 که بازخوانی بیشتری نشان می‌دهند. دقت کلی (Overall Accuracy): دقت کلی مدل با هسته RBF برابر با ۰,۳۹ (۳۹٪) است که بهتر از دقت مدل با هسته خطی (۰,۳۳ یا ۳۳٪) است. ماتریس درهم‌ریختگی: ماتریس درهم‌ریختگی نیز نشان‌دهنده این است که مدل با هسته RBF به طور کلی بهتر عمل می‌کند و تعداد بیشتری از نمونه‌ها را به درستی طبقه‌بندی کرده است. نتیجه‌گیری: مدل SVM با هسته RBF در طبقه‌بندی دستگاه‌های موسیقی عملکرد بهتری نسبت به هسته خطی دارد. این نتیجه ممکن است به دلیل توانایی بیشتر هسته RBF در مدل‌سازی روابط غیرخطی در داده‌ها باشد. برای بهبود بیشتر عملکرد، می‌توان پارامترهای مدل را بهینه‌سازی کرد یا از تکنیک‌های پیش‌پردازش و استخراج ویژگی‌های پیشرفته‌تر استفاده کرد.

(ج)

دقت مدل برای هر کلاس متفاوت است و بالاترین دقت مربوط به کلاس D_6 با مقدار ۰,۵۵ است که نشان می‌دهد مدل در تشخیص نمونه‌های مثبت این کلاس عملکرد بهتری دارد. بازخوانی مدل نیز برای هر کلاس متفاوت است و کلاس D_6 با بازخوانی ۰,۶۵ بهترین عملکرد را داشته است، به این معنا که مدل ۶۵٪ از نمونه‌های واقعی کلاس D_6 را به درستی تشخیص داده است. امتیاز f_1 ، که ترکیبی از دقت و بازخوانی است، برای کلاس D_6 با مقدار ۰,۶۰ بالاترین مقدار را دارد، که نشان‌دهنده تعادل خوبی بین دقت و بازخوانی در این کلاس است.

دقت کلی مدل ۰,۴۱ (۴۱٪) است که بیانگر عملکرد متوسط مدل در طبقه‌بندی کلیه نمونه‌ها است. میانگین‌های ماکرو و وزن‌دار نیز برابر با ۰,۴۱ هستند، که نشان می‌دهد مدل به طور کلی در طبقه‌بندی داده‌ها عملکرد یکنواختی دارد. مدل SVM با هسته RBF و پارامترهای بهینه‌شده عملکرد نسبتاً خوبی در طبقه‌بندی دستگاه‌های موسیقی دارد، به ویژه در تشخیص کلاس D_6.

با این حال، دقت کلی مدل ۴۱٪ است که نشان می‌دهد مدل می‌تواند با بهینه‌سازی بیشتر پارامترها و استفاده از ویژگی‌های بهتر بهبود یابد. پیشنهاد می‌شود برای بهبود بیشتر عملکرد، پارامترهای مدل را بهینه‌تر کرده و از تکنیک‌های پیش‌پردازش و استخراج ویژگی‌های پیشرفته‌تر استفاده شود.

(د)

توضیح Dimention Reduction : متد های PCA – LDA – t-SNE

کاهش ابعاد یکی از مفاهیم کلیدی در پردازش داده‌ها و یادگیری ماشین است. این تکنیک به ما کمک می‌کند تا داده‌های بسیار بزرگ و پیچیده با تعداد فیچر زیاد را به فرمتی ساده‌تر و کوچک‌تر تبدیل کنیم که همچنان اطلاعات مهم را حفظ کرده باشد. این روش برای تجزیه و تحلیل داده‌ها، دسته‌بندی و حتی بصری‌سازی داده‌ها بسیار مفید است. در کل دو هدف از این کار داریم، (۱) representation که در واقع یک دید کلی از داده‌ها به ما می‌دهد و تبدیل می‌کند به فضایی با ابعاد کم مثلاً دو یا سه بعدی؛ و (۲) برای بهبود نتیجه و ساده کردن مدل. از طریق دو روش کلی هم این dimention reduction صورت می‌گیرد: (۱) feature selection که از همون فیچر هامون یه سری هارو استخراج می‌کند و (۲) feature extractions که این مورد دوم سه تا متد اصلی دارد که عبارتند از: تحلیل مؤلفه‌های اصلی (PCA)، تحلیل تمایز خطی (LDA)، و t-Distributed Stochastic Neighbor Embedding (t-SNE). در ادامه به توضیح مختصری درباره هر یک از آن‌ها می‌پردازیم.

تحلیل مؤلفه‌های اصلی (PCA)

PCA یک تکنیک آماری است که برای کاهش ابعاد داده‌ها استفاده می‌شود و این کار را با حفظ مؤلفه‌هایی که بیشترین واریانس را دارند، انجام می‌دهد. اگر X ماتریس داده‌ها باشد، PCA به دنبال مؤلفه‌های اصلی W است که ماکزیمم واریانس را دارا باشند. $XW = Y$ که در آن Y داده‌های کاهش‌یافته است.

در واقع PCA میاد principal component بدست میاره، حالا این چیه؟ یک سری axis هایی که در فضای ویژگی برهم عمودند و در راستای بیشترین واریانس قرار دارند و برهم عمودند. طبق PCA فیچری مهم است که بیشترین واریانس ممکن را دارد و فیچر مهم دوم، فیچری است که بر قبلی عمود است و بر راستای بیشترین واریانس ممکن است و این روال تا آخر ادامه دارد. (تا زمانی که به تعداد ابعاد اصلی دیتاست برسد) و PCA یک

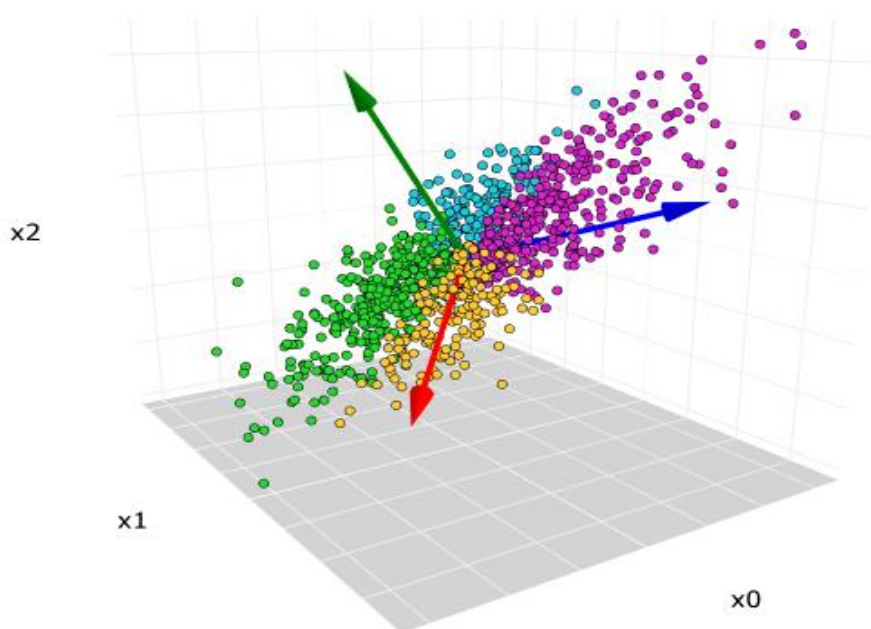
تبدیل خطی است و فیچر هایی که در PCA بدست میاد از هم مستقل هستند که این باعث میشود از over fitting هم جلوگیری شود. (برای unsupervised ها)

حالا چرا بیشترین واریانس برای ما مهمه ؟ در ماشین لرنینگ، واریانس معیاری برای سنجش پراکندگی داده‌ها است و داده‌هایی با بیشترین واریانس بیشترین تغییرات را نشان می‌دهند. این مفهوم در PCA و کاهش ابعاد اهمیت زیادی دارد، زیرا PCA با یافتن ابعادی که بیشترین واریانس را دارند، سعی در حفظ حداکثر اطلاعات با کاهش ابعاد داده‌ها دارد. مراحل PCA شامل استانداردسازی داده‌ها، محاسبه ماتریس کواریانس، و انتخاب مؤلفه‌های اصلی بر اساس بیشترین مقادیر ویژه است. داده‌هایی با واریانس بالا معمولاً اطلاعات بیشتری دارند و می‌توانند به کاهش ابعاد و نویز کمک کنند. در نتیجه، داده‌های با واریانس بالا برای بهینه‌سازی مدل‌ها بسیار مهم هستند.

مزایا: کاهش ابعاد داده‌ها ضمن حفظ اطلاعات و سهولت در بصری‌سازی داده‌های پیچیده.

معایب: حساسیت به مقیاس داده‌ها و نادیده گرفتن وابستگی‌های غیرخطی بین متغیرها.

تحلیل مؤلفه‌های اصلی (PCA) به دلیل ویژگی‌هایی مانند ترکیب خطی مبهم، عدم توجه به معنای ویژگی‌ها، وابستگی به مقیاس داده‌ها و پیچیدگی در تفسیر بارگذاری‌ها، به عنوان یک روش تفسیرپذیر شناخته نمی‌شود. این روش اگرچه برای کاهش ابعاد و شناسایی ساختارهای پنهان در داده‌ها قدرتمند است، اما نتایج آن به سختی قابل تفسیر هستند و معمولاً به عنوان مرحله مقدماتی برای تحلیل‌های بیشتر استفاده می‌شود.



در شکل بالا نمایشی از داده‌های چند بعدی است که با استفاده از تکنیک PCA به فضای سه‌بعدی منتقل شده‌اند. محورها مولفه‌های اصلی را نشان می‌دهند که بیشترین واریانس داده‌ها را توضیح می‌دهند. نقاط رنگی نمایانگر نمونه‌های داده در فضای جدید هستند و فلش‌ها جهت‌های اصلی تغییرات را نشان می‌دهند. این نوع نمایش برای تحلیل و بصری‌سازی داده‌های پیچیده و کشف الگوها و خوشه‌ها در داده‌ها مفید است.

تحلیل تمایز خطی (LDA)

LDA بر خلاف PCA، یک تکنیک نظارت شده است که به دنبال ماکزیمم کردن جدایی بین دسته‌های مختلف داده‌ها است. در LDA، دو ماتریس S_B و S_W تعریف می‌شوند که به ترتیب نشان‌دهنده واریانس داخل کلاس و بین کلاس‌ها هستند. هدف این است که بیشترین نسبت
$$\frac{|W^T S_B W|}{|W^T S_W W|}$$
 بدست آوریم.

LDA برای classification و supervise خیلی عالی و در کل فیچر هایی را درمیاره که در آن ها نمونه های یک کلاس بهم نزدیک اند ولی میانگین دو کلاس متفاوت از هم فاصله داشته باشد یعنی واریانس هر کلاس min و فاصله بین دو کلاس max باشد.

مزایا: بهبود عملکرد دسته‌بندی و مناسب برای برجسته کردن تفاوت‌های بین دسته‌های مختلف.

معایب: کارایی کمتر در داده‌هایی که توزیع نرمال ندارند و محدود به تعداد دسته‌ها کمتر از تعداد ویژگی‌ها.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-SNE یک تکنیک کاهش ابعاد است که برای بصری‌سازی داده‌های دارای ابعاد بالا طراحی شده است (برای representation). این متد با حفظ همسایگی‌های محلی عمل می‌کند. t-SNE ابتدا فواصل بین نقاط در فضای اصلی را به احتمالات تبدیل می‌کند و سپس یک فضای کاهش‌یافته را ایجاد می‌کند که همین احتمالات را تقریباً حفظ می‌کند.

در واقع این متد یک تبدیل غیر خطی می‌زند به نمونه‌ها به قصد پیدا کردن الگو و شباهت بین آن‌ها و نمونه‌های شبیه به هم را به هم نزدیک نگه می‌دارد و نمونه‌هایی که بهم شبیه نیستند را از هم دور نگه می‌دارد و خوب طبیعتاً اولین معیار برای این شباهت هم فاصله اقلیدسی است و هر چه فاصله بیشتر باشد بدون شباهت تریم و

برعکس. این متد از کرنل گوسین استفاده می کند تا پراکندگی دیتاها را از بین ببرد. (برای unsupervised ها)

کرنل گوسین هم به صورت زیر است :

$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x} - \vec{l}^i\|^2}{2\sigma^2}}$$

مزایا: عالی برای بصری سازی دسته های داده ای و حفظ ساختارهای محلی.

معایب: حساس به پارامترهای تنظیم شده و محاسبات نسبتاً سنگین و کند.

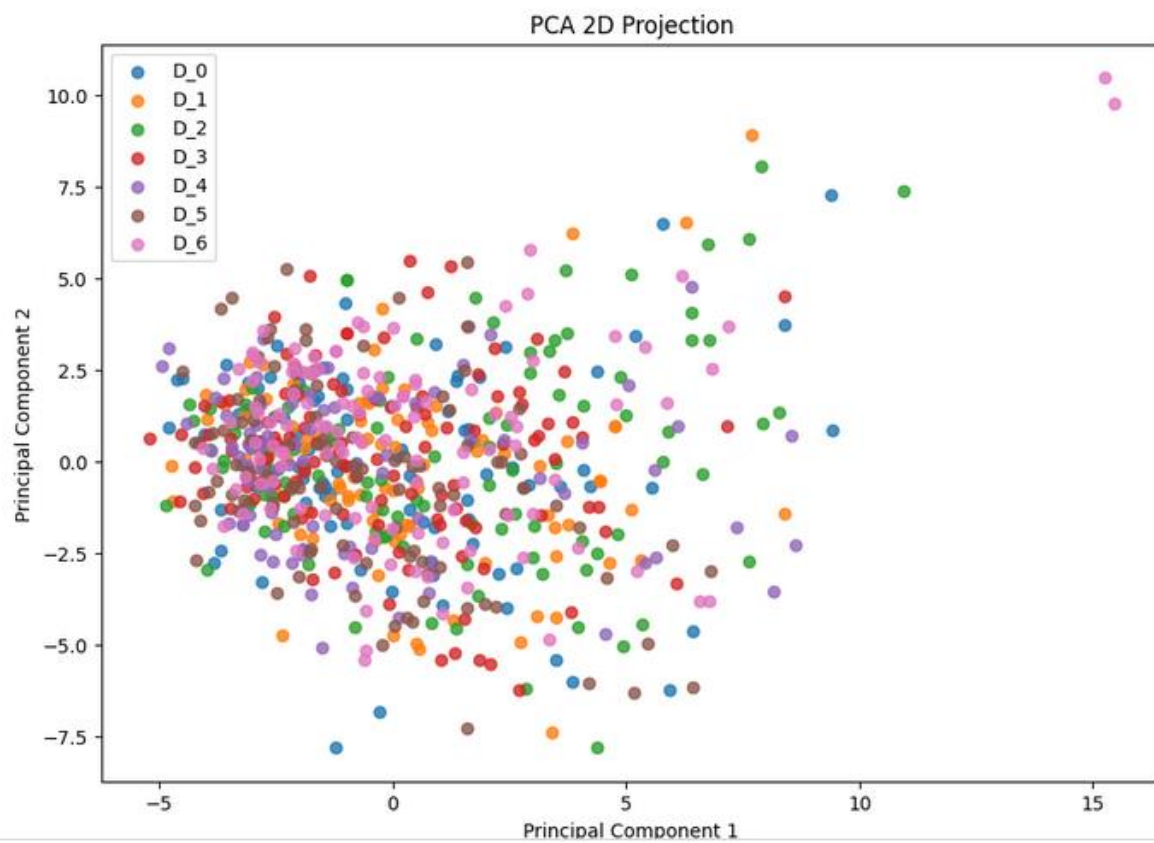
کاربرد این سه متد و یه مقایسه کلی :

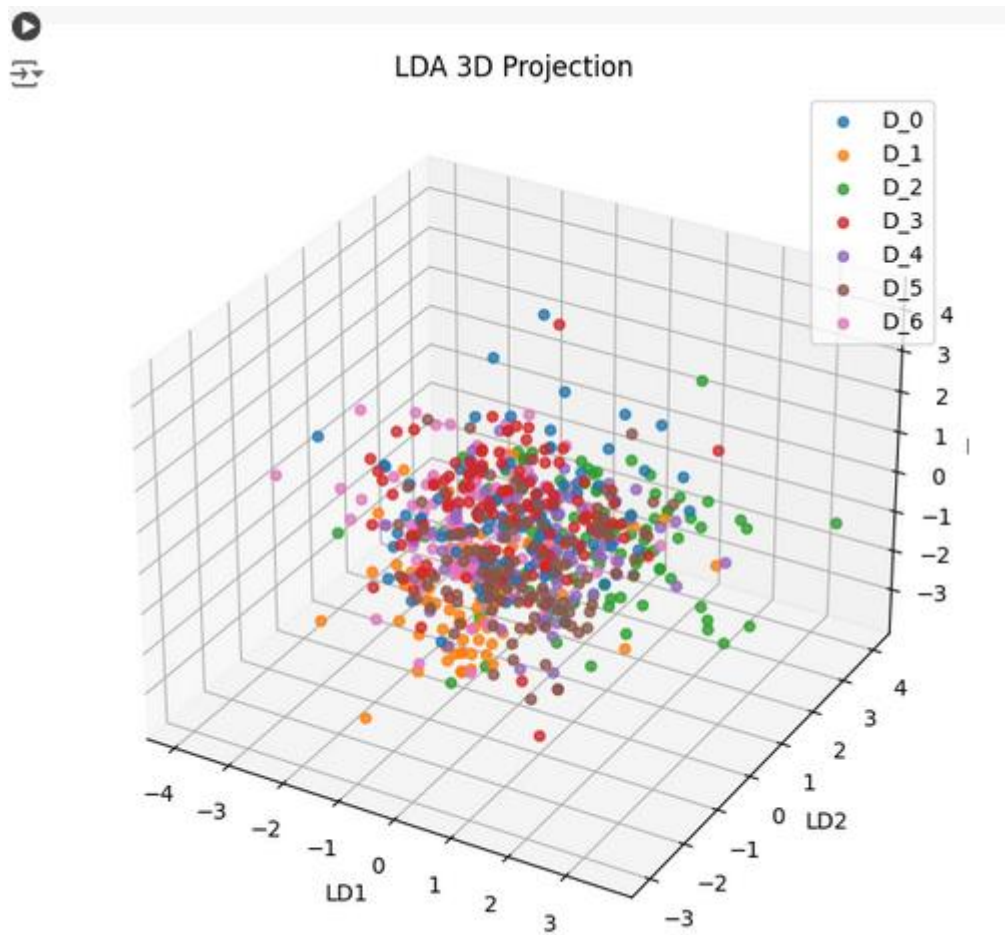
PCA معمولاً در پیش پردازش داده ها برای الگوریتم های یادگیری ماشین به کار می رود و unsupervised است و به دیتای زیادی احتیاج ندارد و برای هم representation و هم کاهش بعد برای تحلیل و بهبود نتایج استفاده میشود.

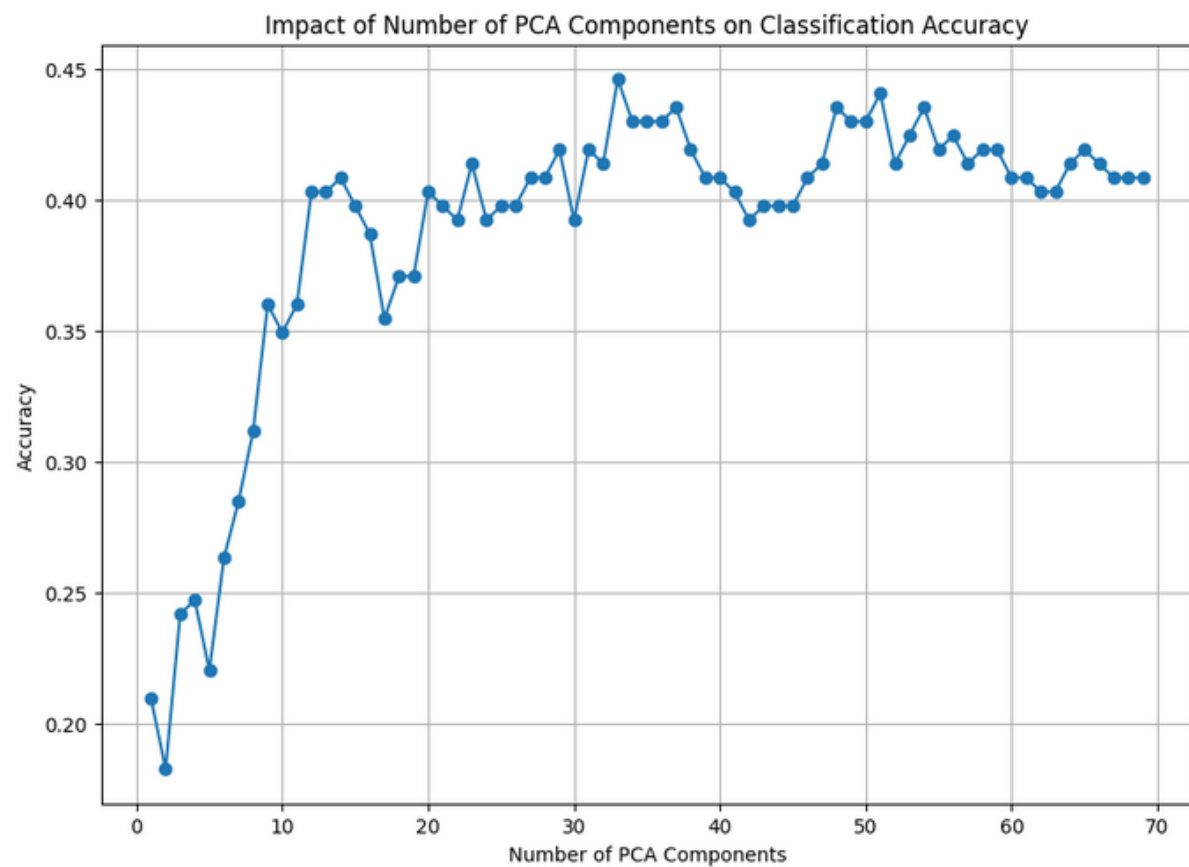
LDA اغلب در موقعیت هایی که داده هایی با برچسب های مشخص هستند استفاده می شود، مانند دسته بندی های تشخیصی یعنی supervise هستند و برای کار classification خیلی خوبه و در دیتاست های بزرگ هم استفاده میشود اصولاً و وقتی که می خواهیم اطلاعات کلاس را مگه داریم یا شبیه نرمال پخش شدن مثل قد افراد، خیلی مناسب است و برای representation مناسب نیستش.

t-SNE برای بصری سازی داده ها در فضای دو یا سه بعدی به کار می رود، به ویژه در تحلیل داده های ژنومیک و سایر داده های پیچیده و برای unsupervised ها و برای کار representation خیلی خوبه.

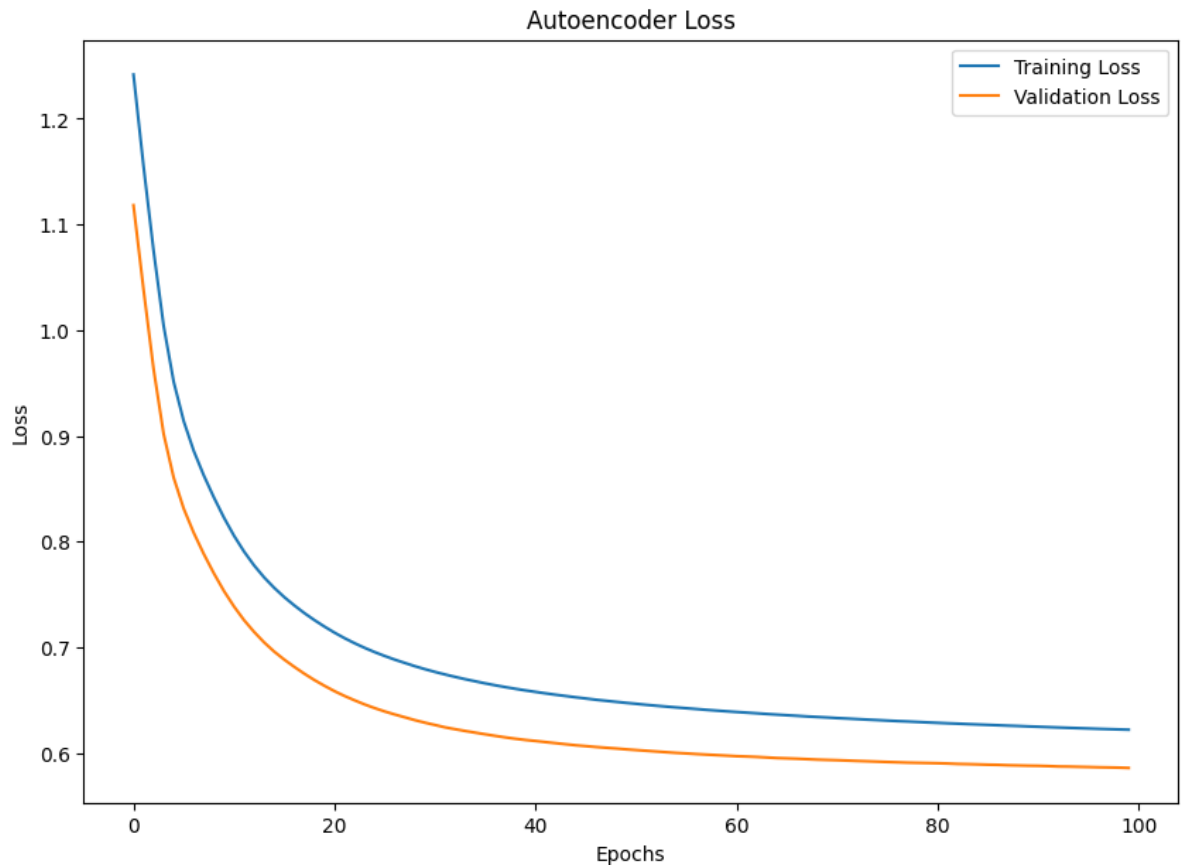
در کل این توضیحات یک نگاه کلی به سه تکنیک مهم کاهش ابعاد بود که هر کدام ویژگی های خاص خود را دارند و بسته به نوع داده و هدف تحلیل، می توان از آنها استفاده کرد.





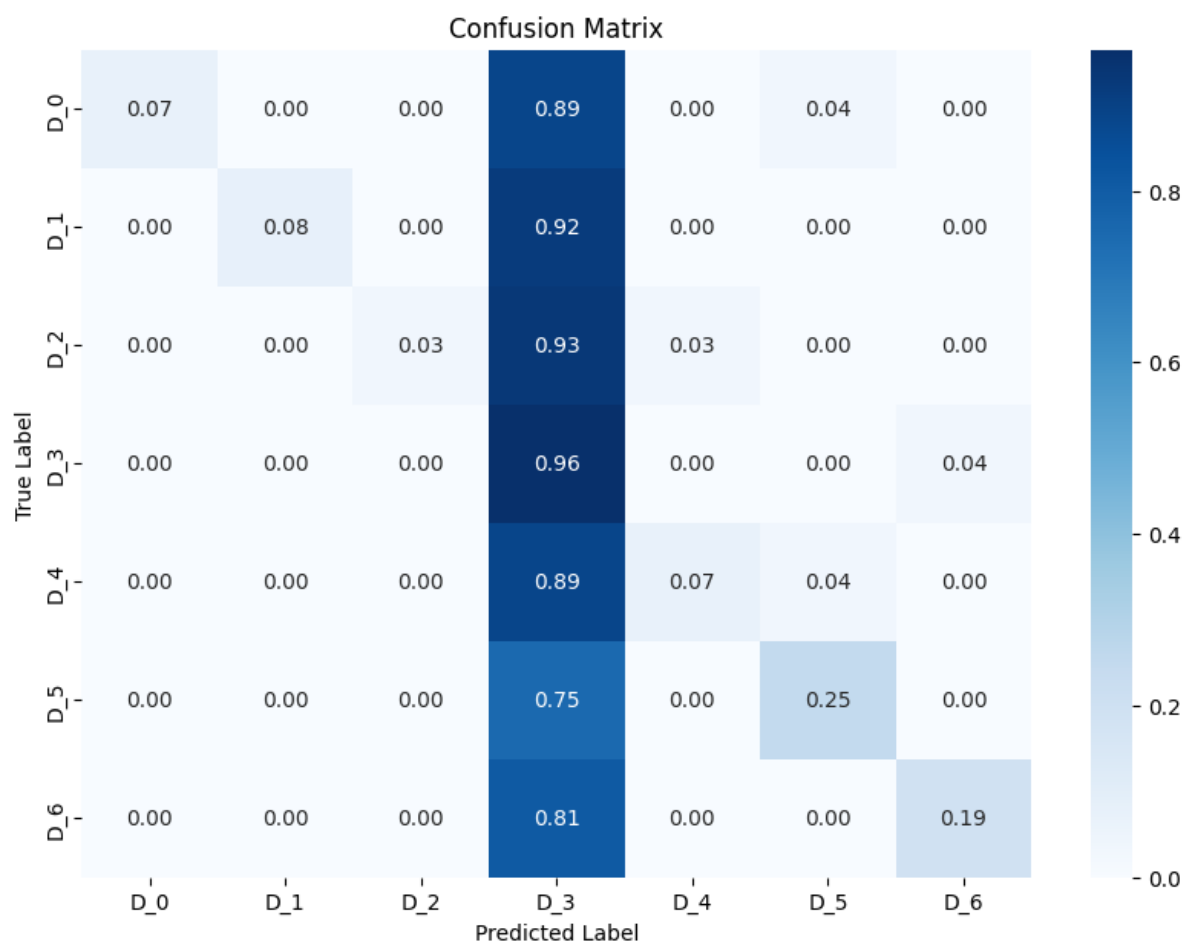


(o

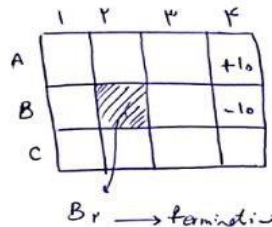


نمودار اتلاف مربوط به بخش آموزش و اعتبارسنجی اتوانکودر نشان‌دهنده روند کاهش خطا در طول دوره‌های آموزشی است. در مراحل اولیه آموزش (تا حدود ۲۰ دوره)، هر دو منحنی خطای آموزش و اعتبارسنجی به سرعت کاهش می‌یابند، که بیانگر یادگیری ویژگی‌های کلیدی از داده‌هاست. کاهش سریع خطا در این مراحل نشان می‌دهد که مدل به خوبی توانسته الگوهای ساده و ابتدایی را از داده‌ها استخراج کند. پس از دوره‌های اولیه، کاهش خطا کندتر شده و منحنی‌ها به تدریج همگرا می‌شوند، که نشان‌دهنده بهینه‌سازی ویژگی‌های پیچیده‌تر و جزئی‌تر است. منحنی اعتبارسنجی به طور پیوسته و بدون نوسانات بزرگ کاهش می‌یابد، که نشان می‌دهد مدل دچار بیش‌برازش (Overfitting) نشده و ویژگی‌های مهم را به خوبی یاد گرفته است. این ثبات نشان‌دهنده تعمیم‌پذیری مناسب مدل است و اینکه مدل می‌تواند ویژگی‌های مهم را از داده‌های جدید و ناشناخته نیز استخراج کند. به نظر می‌رسد که منحنی‌ها در حدود ۶۰ دوره به نقطه‌ای می‌رسند که کاهش خطا بسیار کم شده و منحنی‌ها نزدیک به هم قرار می‌گیرند، که نشان می‌دهد مدل به نقطه بهینه خود رسیده و افزایش دوره‌ها تأثیر زیادی بر بهبود مدل ندارد. اتوانکودر توانسته است ویژگی‌های مهم و مفیدی را از داده‌ها استخراج کند. روند کاهش خطا نشان‌دهنده یادگیری مؤثر و تعمیم‌پذیری مناسب مدل است. ثبات منحنی اعتبارسنجی و عدم وجود نوسانات بزرگ نشان‌دهنده تعمیم‌پذیری خوب و عدم بیش‌برازش مدل است. تعداد دوره‌های بهینه برای آموزش این

اتوانکودر در حدود ۶۰ دوره به نظر می‌رسد، که بعد از آن تغییرات بیشتر تأثیر زیادی بر بهبود مدل ندارد. این نتایج نشان می‌دهد که اتوانکودر به خوبی توانسته ویژگی‌های مهم را از داده‌های صوتی استخراج کند و به بهبود دقت مدل طبقه‌بندی کمک کند.



پرسش ۳



action حالت $\rightarrow A = \{ \text{Left, right, up, down} \}$
که ۴ حرکت ممکن است
 $\begin{matrix} L & r & U & d \end{matrix}$

حالا State ها!

شخص می بیند

	U	d	L	r
A ₁	-	-	-	-
A ₂	-	B ₁	-	A ₂
A ₃	-	B ₂	B ₁	A ₃
A ₄	-	B ₃	A ₂	A ₄
B ₁				
B ₂				
B ₃				
B ₄				
C ₁				
C ₂				
C ₃				
C ₄				

مطابق ماتریس State باید ماتریس Reward

را هم شکل بدهم، پس داریم:

Reward =

	U	d	L	r
A ₁				
A ₂				
A ₃				
A ₄				
B ₁				
B ₂				
B ₃				
B ₄				
C ₁				
C ₂				
C ₃				
C ₄				

با استفاده از مدل Q-learning که رابطه آن به صورت

زیر است باید بتوانیم تقویر ماتریس Q را آپدیت کنیم و Q اولیه را

بهترین حالت است می گیریم.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha [R(s_t, a_t) + \gamma \max_{a'} Q(s', a') - Q(s_t, a_t)]$$

$$\hat{Q}(s, a) = r + \gamma \max_{a'} Q(s', a') \Rightarrow$$

فرض صورت سوال مقدار لا در صورت سوال ۹۰ فرض شده است حالا به سرعت آپدیت کردن Q می دهیم:

این دو را

$$A_1, A_2, A_3, B_3 \rightarrow$$

$$1) Q(A_1, R) = R(A_1, R) + 0.9 \max_{a'} (A_2, a') = 0 \leftarrow \text{episode 1}$$

$$2) Q(A_2, R) = R(A_2, R) + 0.9 \max_{a'} (A_4, a') = 9$$

$$3) Q(A_3, d) = R(A_3, d) + 0.9 \max_{a'} (B_3, a') = 0$$

$$4) Q(B_3, R) = R(B_3, R) + 0.9 \max_{a'} (B_4, a') = -1$$

Q آپدیت اولیہ ماتریس
for episode 1 $\rightarrow Q = \begin{matrix} & \begin{matrix} U & D & L & R \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{matrix} & \begin{bmatrix} - & 0 & - & 1 \\ 0 & 0 & 0 & 9 \\ - & 0 & 0 & 1 \\ - & 0 & 0 & 1 \\ 0 & - & - & - \\ 0 & - & - & - \\ 0 & - & 0 & 0 \\ 0 & - & 0 & 0 \\ 0 & - & 0 & - \\ 0 & - & 0 & - \\ 0 & - & 0 & - \end{bmatrix} \end{matrix}$

episode 2 $\rightarrow \boxed{C_2, C_1, B_2, A_2, A_3, A_4, A_1}$

1) $Q(C_2, L) = R(C_2, L) + 0.9 \max_a Q(C_2, a') = 0$

2) $Q(C_1, U) = R(C_1, U) + 0.9 \max_a Q(B_2, a') = 0$

3) $Q(B_2, U) = R(B_2, U) + 0.9 \max_a Q(A_2, a') = 0$

4) $Q(A_2, R) = R(A_2, R) + 0.9 \max_a Q(A_3, a') = 1.1$

5) $Q(A_3, R) = R(A_3, R) + 0.9 \max_a Q(A_4, a') = 9$

6) $Q(A_4, R) = R(A_4, R) + 0.9 \max_a Q(A_1, a') = 10$

episode 3 $\rightarrow \boxed{C_3, C_2, B_3, A_3, A_4}$

1) $Q(C_3, L) = R(C_3, L) + 0.9 \max_a Q(C_2, a') = 0$

2) $Q(C_2, U) = R(C_2, U) + 0.9 \max_a Q(B_3, a') = 0$

3) $Q(B_3, U) = R(B_3, U) + 0.9 \max_a Q(A_3, a') = 10$

4) $Q(A_3, R) = R(A_3, R) + 0.9 \max_a Q(A_4, a') = 10$

حالا Q نهایی حاصل از هر 3 اپیزود به صورت زیر است (آپدیت)

$Q^* = \begin{matrix} & \begin{matrix} U & D & L & R \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{matrix} & \begin{bmatrix} - & 0 & - & 1 \\ 0 & 0 & 0 & 9 \\ - & 0 & 0 & 1 \\ - & 0 & 0 & 1 \\ 0 & - & - & - \\ 0 & - & - & - \\ 0 & - & 0 & 0 \\ 0 & - & 0 & 0 \\ 0 & - & 0 & - \\ 0 & - & 0 & - \\ 0 & - & 0 & - \end{bmatrix} \end{matrix}$

ماتریس Q^* به صورت روبه راست که هنوز
optimal نیست چون هنوز آپدیت یا به عبارتی
ماپشن نیستیم که آیا مقصودش اینها.

(ب) حال این سیاست را برای π در نظر می گیریم در هر مورد مشخص می کنیم.

	۱	۲	۳	۴
A	→	→	→	↖
B	↓	■	↑	↔
C	↗	↔	↔	↔

این قسمت باید \uparrow باشد ولی نیست

با اقبال $\frac{1}{3}$ برای هر action و Transit است

در نتیجه سیاست π بجهت نیست

این سیاست یک سیاست بجهت بود مثلاً بیان B به صورت \uparrow می بود در ماتریس Q آپدیت شده B را
 میگویند نیست. در اینرد بجهت و item بیشترین update کنیم تا به سیاست بجهت نزدیک تر بشیم.