به نام خدا

تمرین درس یادگیری آماری

سری سوم

امیدرضا داودنیا

زمستان 1402

۱- فرض کنید سه کلاس با توزیع های زیر داریم:

$$P(x|y_1) = N\left(\begin{bmatrix}0.5\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) \qquad P(y_1) = 0.3$$

$$P(x|y_2) = N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}2 & 0.5\\0.5 & 2\end{bmatrix}\right) \qquad P(y_2) = 0.3$$

$$P(x|y_3) = N\left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}5 & 4\\4 & 5\end{bmatrix}\right) \qquad P(y_3) = 0.4$$

احتمال قرار گرفتن نمونه $x^{(0)} = \begin{bmatrix}0.5\\0.5\end{bmatrix}$ در سه توزیع را محاسبه کنید و کلاس این نمونه را مشخص کنید.

طبق قاعده بیز خواهیم داشت ؛ $p(y_i|x^{(0)}) = \frac{p(x^{(0)}|y_i)p(y_i)}{p(x^{(0)})}$ چون مخرج ثابت است هر کلاسی که صورت قائده بیز مربوطه اش بزرگتر باشد کلاس بهتری برای طبقه بندی نمونه صفرم است.

$$p\left(x^{(0)}|y_i\right) = \frac{e^{\frac{-1}{2}(x^{(0)}-\mu_i)^T\Sigma^{-1}(x^{(0)}-\mu_i)}}{2\pi\sqrt{|\Sigma|}}$$

بامحاسبه احتمال های بالا خواهیم داشت ؛

$p(y_1|x^{(0)})p(x^{(0)})$ = 0.1404 * 0.3 = 0.0421

$p(y_2|x^{(0)})p(x^{(0)})$ = 0.0743 * 0.3 = 0.0223

$p(y_3|x^{(0)})p(x^{(0)})$ = 0.0515 * 0.4 = 0.0206

به نظر می رسد بهترین کلاس برای طبقه بندی کلاس اولاست.

۲- تابع توزیع احتمال زیر را در نظر بگیرید:

$$P_\theta(x) = 2\theta x e^{-\theta x^2}$$

که θ پارامتر مسئله و x یک عدد ثابت مثبت است. فرض کنید m نمونه ی x_i که $i.i.d.$ هستند را از این توزیع در اختیار دارید. تخمین $Maximum\ Likelihood$ را برای θ بر اساس نمونه ها بدست آورید.

پس تابع Likelihood بدین صورت می تواند تعریف گردد،به علت استقلال خطی که مسئله داده به راحتی می نویسیم.

$$L(\theta) = \prod_{i=1}^{m} P_\theta(x_i) = \prod_{i=1}^{m} 2\theta x_i e^{-\theta {x_i}^2} = 2\theta \prod_{i=1}^{m} x_i e^{-\theta {x_i}^2}$$

برای پیدا کردن مشتق و ماکزیمم این تابع برای مشتقگیری ساده تر لگاریتم می گیریم.

$$\log L(\theta) = \log\left(2\theta \prod_{i=1}^{m} x_i e^{-\theta {x_i}^2}\right)$$

$$= \log(2) + \log(\theta) + \sum_{i=1}^{m} (\log(x_i) - \theta {x_i}^2)$$

برای پیدا کردن نقطه ماکزیمم با خطی بودن تابع نسبت به پارامتر تتا مشتقگیری میکنیم.

$$\frac{\partial \log L(\theta)}{\partial \theta} = \frac{1}{\theta} - \sum_{i=1}^{m} {x_i}^2 \triangleq 0$$

$$\hat{\theta}_{MLE} = \frac{1}{\sum_{i=1}^{m} {x_i}^2}$$

۳- فرض کنید دیتاستی را جمع آوری کرده‌ایم که در آن بر اساس میزان ساعت مطالعه (X_1) و معدل (X_2)، شانس گرفتن نمره A (Y) را تخمین می‌زنیم. برای رسیدن به این تخمین از یک مدل $logistic\ regression$ استفاده می‌کنیم. پس از کردن fit مدل بر روی داده ها، ضرایب این مدل به صورت $\hat{\beta}_0 = -6, \ \hat{\beta}_1 = 0.05, \ \ \hat{\beta}_2 = 1$ هستند.

الف: احتمال اینکه دانش آموزی با ۴۰ ساعت مطالعه و معدل ۳,۵ بتواند نمره A بگیرد، را محاسبه کنید.

با در نظر گرفتن مسئله به عنوان یم طبقه‌بند لوجیستیک ۲ ورودی ، تک کلاسه می توانیم بنویسیم.

$$t = -6 + 0.05\, x_1 + x_2$$

$$P(Y = A) = \frac{1}{1 + e^{-t}}$$

$$t(40\,.\ 3.5) = -0.5$$

$P(Y = A) = \frac{1}{1+e^{-0.5}} = 0.6224$

ب: این دانش آموز با معدل ۳٫۵، برای اینکه50% شانس دریافت نمره A را داشته باشد، باید حداقل چند ساعت مطالعه داشته باشد؟

$$P(Y = A) = \frac{1}{1+e^{-t}} \triangleq 0.5$$

$$e^{-t} = 1 \ \rightarrow t = \ -6 + 0.05\, x_1 + \ 3.5 = 0$$

$$x_1^{\%50} = 50\ hours$$

۴- یک مسئله طبقه بندی دو کلاسه را با نمونه های شامل یک ویژگی در نظر بگیرید. نمونه های کلاس ۱ به طور یکنواخت در بازه ی $[-5,5]$ و نمونه های کلاس صفر به طور یکنواخت در بازه $[a,b]$ قرار دارند. می‌خواهیم مرز طبقه بندی با روش تحلیل افتراقی خطی (LDA) پیدا کنیم. توجه داشته باشید که در این روش، واریانس نمونه های دو کلاس، برابر فرض می‌شوند.

اگر فرض کنیم که احتمال های اولیه دو کلاس با هم برابر باشند، آیا در این مسئله مقدار واریانس نمونه ها، در محل قرارگیری مرز طبقه بندی موثر است؟ جواب خود را اثبات کنید.

$$\mathrm{P(X = x|Y = 1)P(Y = 1) = P(X = x|Y = 0)P(Y = 0)}$$

اگر احتمال اولیه ها برابر باشند خواهیم داشت.

$$\mathrm{P(X = x|Y = 1) = P(X = x|Y = 0)}$$

تابع چگالی احتمال توزیع یکنواخت پیوسته بدین صورت است ؛

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & else \end{cases} = \frac{1}{b-a}(u(x-a) - u(x-b))$$

واریانس این توزیع بدین صورت محاسبه می گردد ؛

$\mathrm{Var}(x)=\frac{(b-a)^2}{12}$

که در کیس ما اگر واریانس ها برابر باشند بدین معناست که می بایست عرض هر دو توزیع ۱۰ باشد.

$$\frac{1}{b-a}\big(u(x-a)-u(x-b)\big)=\frac{1}{10}\big(u(x+5)-u(x-5)\big)$$

اگر واریانس ها را برابر فرض کنیم در این صورت عرض هر دو توزیع نیز باید برابر باشد پس خواهیم داشت.

در این حالت که واریانس هر دو توزیع برابر است مرز طبقه بندی بین ۲ میانگین رخ خواهد داد.

ولی در صورتی که یک توزیع از دیگری عریض تر باش واریانس در مرز تصمیم گیری موثر خواهد بود.

$$w = \Sigma^{-1}(\mu_a-\mu_b)$$

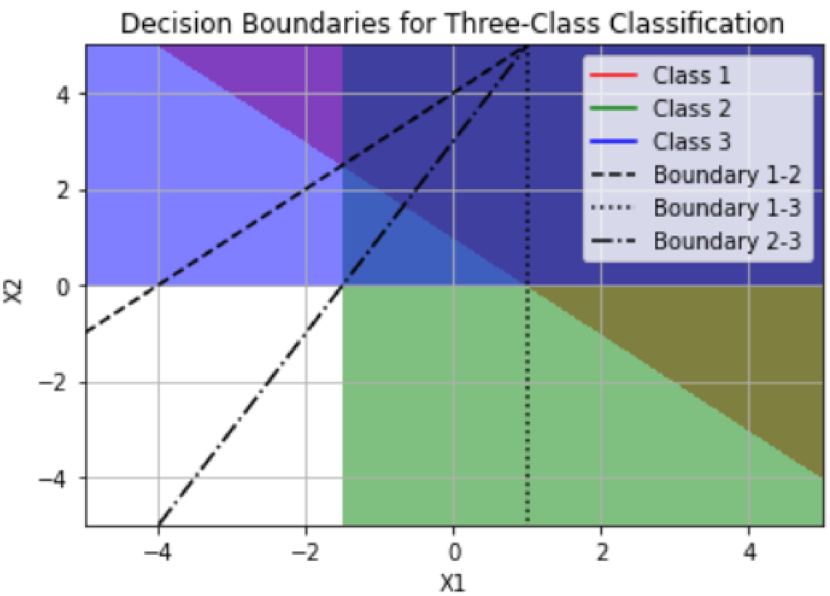
۵- یک مسئله طبقه بندی سه کلاسه را با دو ویژگی (X_2, X_1) در نظر بگیرید. فرض کنید *Discriminant Function* هر کلاس را مطابق زیر بدست آورده ایم:

$$\delta_1(x) = x_1 + x_2 - 1$$

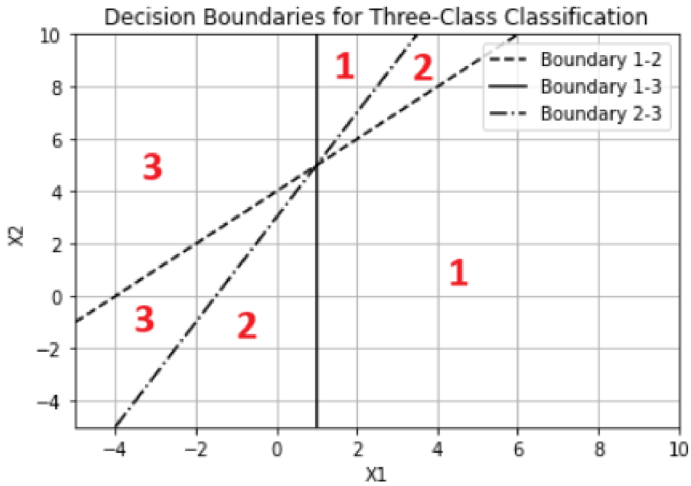
$$\delta_2(x) = 2x_1 + 3$$

$$\delta_3(x) = x_2$$

نواحی را در فضای $\mathbb{R}^2$ که هر کلاس به آن تعلق دارد را رسم کنید.



برای رسم این نواحی هر بار تابع مرز تصمیم گیری ها مقایسه شده و خطوط رسم شده اند و همچنین ، نواحی رنگی مربوط به بزرگ تر از صفر بودن نواحی تصمیم گیری است.



۶ - فرض کنید π_A , π_B نشان دهنده $P(Y=B)$ و $P(Y=A)$ باشند و (μ_B,Σ_B) و (μ_A,Σ_A) میانگین و ماتریس کواریانس کلاس های A و B را مشخص کنند.

$$N(\mu_A,\Sigma_A)=\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}2 & 0.5\\0.5 & 2\end{bmatrix}\right),\pi_A=0.6$$

$$N(\mu_B,\Sigma_B)=\left(\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}5 & 4\\4 & 5\end{bmatrix}\right),\pi_B=0.4$$

اگر از QDA برای طبقه بندی استفاده کنیم، مرز تصمیم گیری را بدست آورید.

$$\left(-\frac{1}{2}\log|\Sigma_A|-\frac{1}{2}(x-\mu_A)^T\Sigma_A^{-1}(x-\mu_A)+log\pi_A\right)-\left(-\frac{1}{2}\log|\Sigma_B|-\frac{1}{2}(x-\mu_B)^T\Sigma_A^{-1}(x-\mu_B)+log\pi_B\right)=0$$

برای محاسبه ترم $(x-\mu)^T\Sigma^{-1}(x-\mu)$ ابتدا فرض می کنیم بردار x بدین صورت است $\begin{bmatrix}x_1\\x_2\end{bmatrix}$ و معکوس ماتریس کواریانس را فعلا به صورت $\Sigma^{-1}=\begin{bmatrix}a & b\\c & d\end{bmatrix}$ در نظر میگیریم .

در این صورت ؛

$$f(x)=(x-\mu)^T\Sigma^{-1}(x-\mu)=ax_1^2+(b+c)x_1x_2+dx_2^2$$

$$\Sigma_A^{-1}=\begin{bmatrix}0.53 & -0.13\\-0.13 & 0.53\end{bmatrix},\Sigma_B^{-1}=\begin{bmatrix}0.55 & -0.44\\-0.44 & 0.55\end{bmatrix}$$

$$f_A(x) = 0.53x_1^2 - 0.26x_2 + 0.53x_2^2$$

$$f_A(x) = 0.55x_1^2 - 0.88x_1x_2 + 0.55x_2^2$$

$$\left(-0.6608 - \frac{1}{2}(0.53x_1^2 - 0.26x_2 + 0.53x_2^2) - 0.5108\right)$$
$$-\left(-1.0986 - \frac{1}{2}(0.55x_1^2 - 0.88x_1x_2 + 0.55x_2^2) - 0.9162\right) = 0$$

با ساده سازی بیشتر خواهیم داشت.

$$1.6864 + 0.02x_1^2 - 0.62x_1x_2 + 0.02x_2^2 = 0$$

سوالات بخش شبیه سازی به خوبی در فایل نوتبوک پیاده سازی و گزارش شده اند.