بسمه تعالی

دانشکده مهندسی برق و کامپیوتر

تمرین سری اول یادگیری ماشین

محمد ناصری

۸۱۰۱۰۰۴۸۶

Contents

سوال اول 3
سوال دوم 7
سوال سوم 9
سوال چهارم 11
سوال پنجم 13
سوال ششم 14

سوال اول

فرض کنید دو کلاس داده A و B به صورت شکل زیر تقسیم شده اند. داده های کلاس A با رنگ قرمز و داده های دسته B با رنگ آبی مشخص شده اند. با فرض توزیع گاوسی برای هر کلاس، معادله مرز تصمیم مابین این دو کلاس را محاسبه کنید. تمامی مقادیر لازم)میانگین، احتمال پیشین و ...(را از مشاهدات براساس نمودار بدست بیاورید. همچنین مرز بدست آمده از محاسبات را بر روی شکل رسم کنید و درستی جواب خود را بررسی کنید.

نقاط:

(-4, 0), (-3, 1), (-2, -2), (-2, -1), (0, -4)

(-2, 1), (-1, 0), (1, 1), (3,2), (3,5), (4,3), (5,2)

در ابتدا نیاز است تا میانگین یا μ را برای هر دسته حساب کنیم:

$$\mu_1 = \left(\frac{-4-3-2-2+0}{5}, \frac{0+1-2-1-4}{5}\right) = \begin{pmatrix}-2.2,\\ -1.2\end{pmatrix}$$

$$\mu_2 = \left(\frac{-2-1+1+3+3+4+5}{7}, \frac{1+0+1+2+5+3+2}{7}\right) = \begin{pmatrix}1.85\\ 2\end{pmatrix}$$

سپس مقادیر واریانس را محاسبه میکنیم:

$$Var_1(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = 1.76$$

$$Var_1(Y) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2 = 2.96$$

$$Var_2(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = 5.83$$

$$Var_2(Y) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2 = 2.28$$

در مرحله بعد مقادیر کواریانس را محاسبه میکنیم:

$$Cov_1 = \sigma_{XY} = \frac{1}{x}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) = -2.04$$

$$Cov_2 = \sigma_{XY} = \frac{1}{x}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) = 2.28$$

حال ماتریس کواریانس را بدست می آوریم:

$$\Sigma_1 = \begin{pmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{pmatrix} = \begin{bmatrix} 1.76 & -2.04 \\ -2.04 & 2.96 \end{bmatrix}$$

$$\Sigma_2 = \begin{pmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{pmatrix} = \begin{bmatrix} 5.84 & 2.28 \\ 2.28 & 2.28 \end{bmatrix}$$

وارون این ماتریس ها نیز محاسبه میکنیم:

$${\Sigma_1}^{-1} = \begin{bmatrix} 2.82 & 1.94 \\ 1.94 & 1.67 \end{bmatrix}$$

$${\Sigma_2}^{-1} = \begin{bmatrix} 0.28 & -0.28 \\ -0.28 & 0.71 \end{bmatrix}$$

حال احتمال prior ها را بدست می آوریم:

در کل در این مساله ۱۲ نمونه داریم که از این تعداد ۵ عدد از دسته اول و ۷ عدد از دسته دوم هستند پس برای احتمال پیشین میتوانیم بگوییم:

$$P(w_1) = \frac{5}{12}\ , P(w_2) = \frac{7}{12}$$

و بر طبق روابط داریم :

$$g_i(\mathbf{x}) = \mathbf{x}^t \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^t \mathbf{x} + w_{i0},$$

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1},$$

$$\mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^t \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln\,|\boldsymbol{\Sigma}_i| + \ln\,P(\omega_i).$$

برای دسته اول داریم:

$$W_1 = -\frac{1}{2}\begin{bmatrix}2.82 & 1.94\\ 1.94 & 1.67\end{bmatrix} = \begin{bmatrix}-1.41 & -0.97\\ -0.97 & -0.835\end{bmatrix}$$

$$w_1 = \begin{bmatrix}2.82 & 1.94\\ 1.94 & 1.67\end{bmatrix} * \begin{bmatrix}-2.2\\ -1.2\end{bmatrix} = \begin{bmatrix}-8.532\\ -6.272\end{bmatrix}$$

$$w_{10} = -\frac{1}{2} * [-2.2, -1.2] * \begin{bmatrix}2.82 & 1.94\\ 1.94 & 1.67\end{bmatrix} * \begin{bmatrix}-2.2\\ -1.2\end{bmatrix}$$
$$-\frac{1}{2}\ln\left(\left|\begin{bmatrix}2.82 & 1.94\\ 1.94 & 1.67\end{bmatrix}\right|\right) + \ln\left(\frac{5}{12}\right) = (-13.1484) - 0.72 - 0.875$$
$$= -14.7434$$

$$g_1(x) = [x_1, x_2] * \begin{bmatrix}-1.41 & -0.97\\ -0.97 & -0.835\end{bmatrix} * \begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}-8.532\\ -6.272\end{bmatrix} * \begin{bmatrix}x_1\\ x_2\end{bmatrix} - 14.7434$$
$$= -1.41x_1^2 - 1.94x_1x_2 - 0.835x_2^2 - 8.532x_1 - 6.272x_2 - 14.7434$$

برای دسته دوم داریم:

$$W_2 = -\frac{1}{2}\begin{bmatrix}0.28 & -0.28\\ -0.28 & 0.71\end{bmatrix} = \begin{bmatrix}-0.14 & 0.14\\ 0.14 & -0.355\end{bmatrix}$$

$$w_2 = \begin{bmatrix}0.28 & -0.28\\ -0.28 & 0.71\end{bmatrix} * \begin{bmatrix}1.85\\ 2\end{bmatrix} = \begin{bmatrix}-0.042\\ 0.902\end{bmatrix}$$

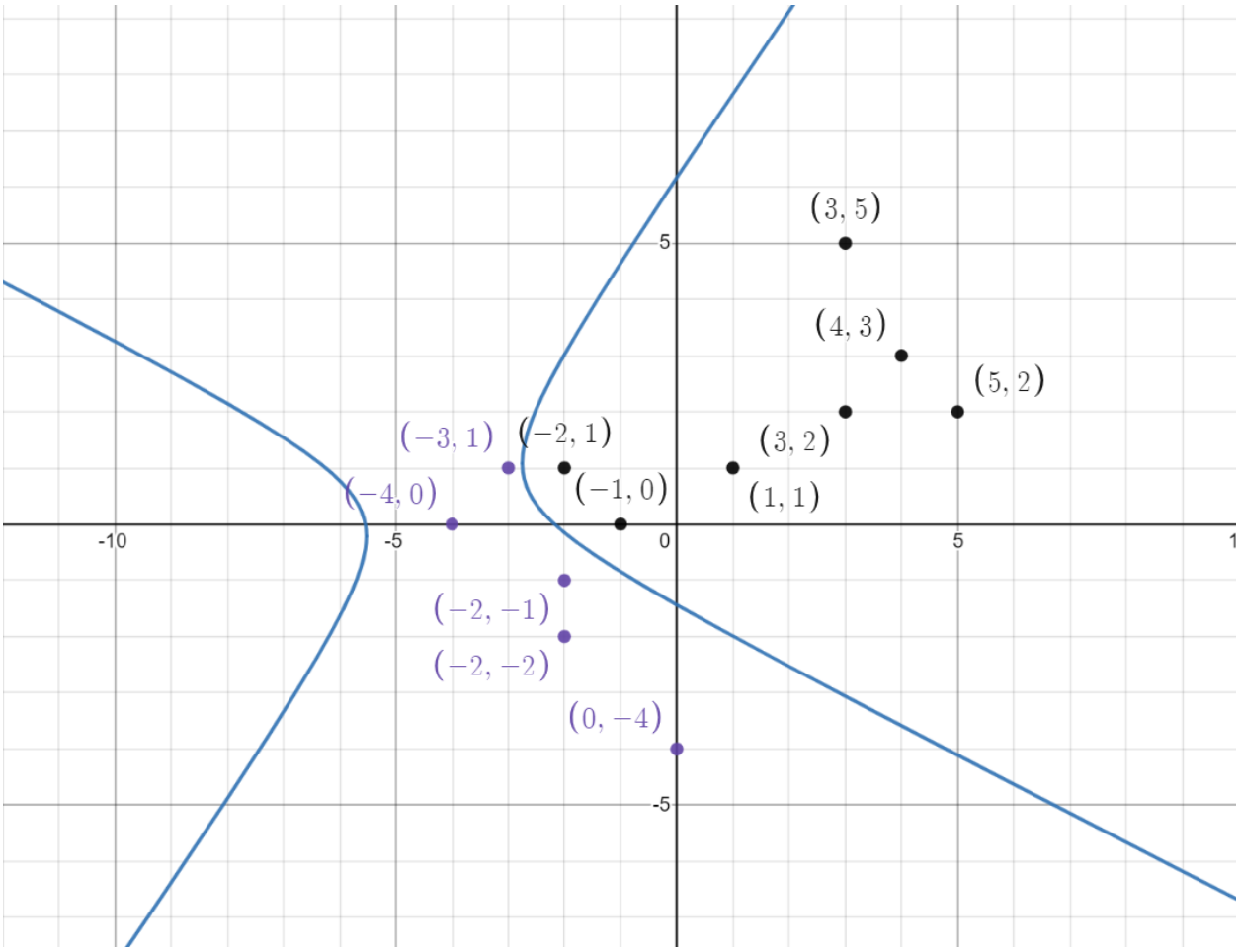
$$w_{20} = -\frac{1}{2} * [1.85,2] * \begin{bmatrix}0.28 & -0.28\\ -0.28 & 0.71\end{bmatrix} * \begin{bmatrix}1.85\\ 2\end{bmatrix}$$
$$-\frac{1}{2}\ln\left(\left|\begin{bmatrix}0.28 & -0.28\\ -0.28 & 0.71\end{bmatrix}\right|\right) + \ln\left(\frac{7}{12}\right) = (-0.86315) - 0.0755 - 0.538$$
$$= -1.47665$$

$$g_2(x) = [x_1, x_2] * \begin{bmatrix}0.28 & -0.28\\ -0.28 & 0.71\end{bmatrix} * \begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}-0.042\\ 0.902\end{bmatrix} * \begin{bmatrix}x_1\\ x_2\end{bmatrix} - 1.47665$$
$$= -0.28x_1^2 - 0.56x_1x_2 + 0.71x_2^2 - 0.042x_1 + 0.902x_2 - 1.47665$$

در آخر داریم:

$$g(x) = g(x) \rightarrow -1.41x_1^2 - 1.94x_1x_2 - 0.835x_2^2 - 8.532x_1 - 6.272x_2 - 14.7434$$
$$= -0.28x_1^2 - 0.56x_1x_2 + 0.71x_2^2 - 0.042x_1 + 0.902x_2 - 1.47665$$
$$\Rightarrow -1.1x_1^2 - 1.4x_1x_2 + 1.5x_2^2 - 8.5x_1 - 7.1x_2 - 13.3 = 0$$

با رسم مرز بدست آمده(بدلیل محاسبات دستی، از اعشار جهت سادگی محاسبات صرف نظر شده فلذا امکان خطای کمی وجود دارد) داریم:



مشاهده میشود که مرز به درستی تشکیل و دسته‌ها را از یکدیگر جدا میکند.

سوال دوم

فرض کنید چگالی احتمال برای داده های دو کلاس مختلف به صورت زیر باشد:

$$p(x|\omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_i}{b}\right)^2}, \qquad i = 1, 2.$$

فرض کنید که احتمال پیشین دو کلاس با هم برابر می باشد. (20 نمره)

الف) نشان دهید که حداقل احتمال خطا به صورت رابطه زیر بدست می آید:

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|.$$

احتمال خطا در طبقه بندی باینری به صورت زیر تعریف میشود:

$$P(error) = \begin{cases} p(w_1|x) & if\ we\ decide\ w_2 \\ p(w_2|x) & if\ we\ decide\ w_1 \end{cases}$$

یا به تعبیری دیگر داریم:

$$P(error) = p(x \in R_2, y = 1) + p(x \in R_1, y = 2)$$

$$P(error) = \int_{R_1} p(x|w_2)p(w_2)dx + \int_{R_2} p(x|w_1)p(w_1)dx$$

که از آنجایی که p(x) از w مستقل است میتوان از آن صرف نظر کرد. همچنین با توجه به صورت سوال احتمالات prior این دو کلاس برابر است پس داریم $P(w_i) = 1/2$ با توجه به این تعریف و صورت سوال خواهیم داشت:

$$P(error) = \int_{R_1} \frac{1}{\pi b} * \frac{1}{1 + \left(\frac{x - a_2}{b}\right)^2} * \frac{1}{2} dx + \int_{R_2} \frac{1}{\pi b} * \frac{1}{1 + \left(\frac{x - a_1}{b}\right)^2} * \frac{1}{2} dx$$

از آنجایی که در این مساله احتمال پیشین دو کلاس برابر و همچنین تابع loss نداریم، مرز تصمیم بین دو کلاس در وسط آن دو یعنی $\frac{a_2+a_1}{2}$ در نظر میگیریم. همچنین فرض میکنیم $a_2 > a_1$ حال با توجه به قانون بالا داریم:

$$P(error) = \int_{-\infty}^{\frac{a_2+a_1}{2}} \frac{1}{\pi b} * \frac{\frac{1}{2}}{1 + \left(\frac{x - a_2}{b}\right)^2} dx + \int_{\frac{a_2+a_1}{2}}^{+\infty} \frac{1}{\pi b} * \frac{\frac{1}{2}}{1 + \left(\frac{x - a_1}{b}\right)^2} dx$$

$$= \int_{-\infty}^{\frac{a_2+a_1}{2}} \frac{1}{\pi b} * \frac{\frac{1}{2}}{1+\left(\frac{x-a_2}{b}\right)^2} dx + \int_{\frac{a_2+a_1}{2}}^{+\infty} \frac{1}{\pi b} * \frac{\frac{1}{2}}{1+\left(\frac{x-a_1}{b}\right)^2} dx$$

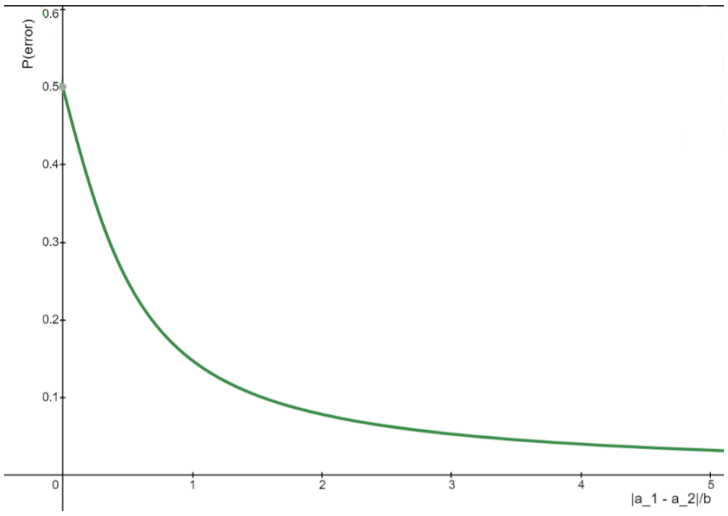
$$= \frac{1}{\pi b} * \int_{-\infty}^{\frac{a_2+a_1}{2}} \frac{\frac{1}{2}}{1+\left(\frac{x-a_2}{b}\right)^2} dx + \frac{1}{\pi b} * \int_{\frac{a_2+a_1}{2}}^{+\infty} \frac{\frac{1}{2}}{1+\left(\frac{x-a_1}{b}\right)^2} dx$$

$$= \frac{1}{\pi b} * \int_{-\infty}^{\frac{a_1-a_2}{2}} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} dx = \frac{1}{\pi b}\left(b.\arctan\left(\frac{-a_1-a_2}{2b}\right) + \frac{\pi b}{2}\right)$$

$$= \frac{2\arctan\left(\frac{-a_1-a_2}{2b}\right)+\pi}{2\pi} = \frac{\pi}{2\pi} - \frac{2\arctan\left(\left|\frac{a_1-a_2}{2b}\right|\right)}{\pi} = \frac{1}{2} - \frac{1}{\pi} tan^{-1}\left|\frac{a_1-a_2}{2b}\right|$$

ب) تابع بدست آمده را به صورت تابعی از $|a2-a1|/b$ رسم کنید.

برای رسم این نمودار نمودار معادله $\frac{1}{2} - \frac{1}{\pi}\arctan(2x)$ را رسم میکنیم.



ج) بیشترین مقدار خطا P(error) چه مقدار می باشد و تحت چه ظرایطی این اتفاق رخ میدهد؟ توضیح دهید.

بیشترین مقدار احتمال خطا برای این مساله برابر 0.5 میباشد که زمانی رخ میدهد که $\left|\frac{a_1-a_2}{2b}\right| = 0$ و این امر زمانی حاصل میشود که مقدار $a_1 = a_2$ باشد یعنی دو توزیع برابر باشند و یا زمانی که مخرج کسر یعنی b به سمت بینهایت میل کند.

سوال سوم

در این سوال میخواهیم برخی تخمینگر های MLEرا بررسی کنیم

الف) یک تاس mوجهی را nبار پرتاب میکنیم. در صورتی که هر وجه xiبار ظاهر شود، تخمینگر MLEاز piها که احتمال آمدن هر وجه می باشد را بیابید

$$f(x_1, \dots, x_m | p_1, \dots, p_m) = \frac{n!}{\pi x_i!} \prod_{i=1}^{m} {p_i}^{x_i} = n! \prod_{i=1}^{m} \frac{{p_i}^{x_i}}{x_i!}$$

$$\log(f) = \log(n!) \sum_{i=1}^{m} x_i log p_i - \sum_{i=1}^{m} x_i!$$

$$L(p, \lambda) = \log\big(L(p)\big) + \lambda(1 - \sum_{i=1}^{m} p_i$$

$$\frac{\partial}{\partial p_i} L(p, \lambda) = \frac{\partial}{\partial p_i} \left(\log(n!) \sum_{i=1}^{m} x_i log p_i - \sum_{i=1}^{m} x_i! \right) - \lambda = \frac{x_i}{p_i} - \lambda$$

با قرار دادن مشتق برابر صفر داریم:

$$\frac{\partial}{\partial p_i} L(p, \lambda) = \frac{x_i}{p_i} - \lambda = 0 \Rightarrow p_i = \frac{x_i}{\lambda} = \frac{x_i}{n}$$

$$\frac{1}{\lambda} \sum_{i=1}^{m} x_i = 1 \Rightarrow \lambda = n$$

ب) متغیر تصادفی Xدارای توزیع Unif([0,θ ([می باشد. با فرض مشاهده nنمونه iidاز این توزیع، تخمینگر MLEازθ را بیابید

$$f(x_i|\theta) = \begin{cases} \frac{1}{\theta} & 0 < x_i < \theta \\ 0 & other \end{cases}$$

$$L(\theta) = \prod f(x_i|\theta) = \frac{1}{\theta^n}$$

از آنجایی که $L(\theta)$ یک تابع نزولی است برای مقدار ماکسیمم، بایستی θ کمترین مقدار را داشته باشد(استفاده از min)

ج) توزیع تخمینگر قسمت ب را بیابید.بررسی کنید که این تخمینگر unbiasedاست یا نه

د)n نمونه iidاز یک توزیع پواسون با پارامتر λداریم. تخمینگر MLEاین پارامتر را بدست آورید

$$\mathrm{p}(x_i|\theta) = \frac{e^{-\theta}\theta^{x_i}}{x_i!}$$

$$L(\theta) = \prod \frac{e^{-\theta}\theta^{x_i}}{x_i!}$$

$$\ln\big(L(\theta)\big) = \sum -\theta \ln(\mathrm{e}) + x_i \ln(\theta) - \ln(x_i)$$

$$= -n\theta + \ln(\theta)\sum x_i - \sum \ln(x_i!)$$

$$\frac{\partial}{\partial\theta}\ln\big(L(\theta)\big) = -n + \frac{1}{\theta}\sum x_i = 0 \Rightarrow \theta = \frac{\sum x_i}{n}$$

سوال چهارم

فرض کنید تعدادی داده از یک توزیع گاوسی با کوواریانس Σ که مشخص است و میانگین μ که اطلاعی از آن نداریم در دست داریم. حال فرض کنید که این میانگین خود یک متغیر تصادفی باشد و از یک توزیع گاوسی با میانگین m_0و کوواریانس Σ_0باشد.

الف) تخمین MAP برای μ را بیابید.

با توجه به توزیع گاوسی داریم:

$$\ln\big(p(D|\mu)\big) = \ln\Big(\prod_{i=1}^{n} p(X_i|\mu)\Big) = \frac{-n\ln((2\pi)^d|\Sigma|)}{2} - \sum_{i=1}^{n}\frac{(X_i-\mu)^T\Sigma^{-1}(X_i-\mu)}{2}$$

$$p(\mu) = \frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}}\exp\Big(\frac{-(\mu-m_0)^T{\Sigma_0}^{-1}(\mu-m_0)}{2}\Big)$$

برای تخمین MAP برای μ داریم:

$$\hat{\mu} = argmax[ln(p(D|\mu)p(\mu))]$$

ب) فرض کنید که مختصات فضا را توسط یک تبدیل خطی به صورت X’= AX ، که در آن A یک ماتریس غیرتکین است.

بررسی کنید که آیا تخمین بدست آمده در قسمت الف، تخمین مناسبی از μ’ نیز می باشد یا نه.

$$\begin{aligned}\mu_2 &= \varepsilon[AX] = A\varepsilon[X] = A\mu\\ \Sigma_2 &= \varepsilon[(Ax_2-\mu_2)(Ax_2-\mu_2)^T] = \varepsilon[A(x_2-\mu_2)(x_2-\mu_2)^TA^T]\\ &= A\varepsilon[(x_2-\mu_2)(x_2-\mu_2)^T]A^T = A\Sigma A^T\end{aligned}$$

$$\begin{aligned}\ln\big(p(D_2|\mu_2) &= \ln\Big(\prod_{i=1}^{n} p(AX_i|A\mu) = \frac{-n\ln((2\pi)^d|A\Sigma A^T|)}{2}\\ &- \sum_{i=1}^{n}\frac{(AX_i-A\mu)^T(A\Sigma A^T)^{-1}(AX_i-A\mu)}{2}\\ &= \frac{-n\ln((2\pi)^d|A\Sigma A^T|)}{2}\\ &- \sum_{i=1}^{n}\frac{((X_i-\mu)^TA^T)((A^{-1})^T\Sigma^{-1}(A^{-1}))(A(X_i-\mu))}{2}\\ &= \frac{-n\ln((2\pi)^d|A\Sigma A^T|)}{2} - \sum_{i=1}^{n}\frac{(X_i-\mu)^T(A^T(A^{-1})^T)\Sigma^{-1}(A^{-1}A)(X_i-\mu)}{2}\end{aligned}$$

از آنجایی که $AA^{-1} = 1$ خواهیم داشت:

$$\ln\left(p(D_2|\mu_2) = \frac{-n\ln((2\pi)^d|A\Sigma A^T|)}{2} - \sum_{i=1}^{n}\frac{(X_i-\mu)^T\Sigma^{-1}(X_i-\mu)}{2}\right.$$

همچنین خواهیم داشت:

$$p(\mu_2) = \frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}}\exp\left(\frac{-(A\mu - Am_0)^T(A\Sigma_0 A^T)^{-1}(A\mu - Am_0)}{2}\right)$$

با همان روش قسمت قبل و رسیدن به AA^{-1} خواهیم داشت:

$$p(\mu_2) = p(\mu) = \frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}}\exp\left(\frac{-(\mu - m_0)^T{\Sigma_0}^{-1}(\mu - m_0)}{2}\right)$$

برای تخمین MAP برای μ_2 داریم:

$$\hat{\mu}_2 = argmax[ln(p(D|\mu)p(\mu))]$$

با مقایسه دو فرم بدست آمده متوجه میشویم که برابر هستند و در نتیجه تخمینگر ما برای مدل تبدیل شده نیز به درستی عمل میکند

سوال پنجم

در این سوال، طبقه بندی طراحی کنید که بتوانیم، که . ۲ کلاس متفاوت) دو تیم فوتبال منچستریونایتد و چلسی(با استفاده از دیتاست داده شده، را تشخیص دهیم. جهت طبقهبندی، میتوانید میانگین رنگ در هر عکس را محاسبه نمایید، سپس بر اساس مقدار به دست آمده، با مقدار رنگ آبی و قرمز مقایسه نمایید. برای دیتاست داده شده، این طبقه بند را تست کنید. ماتریس Confusionرا گزارش دهید. مقادیر accuracy ، precisionو recall را محاسبه کنید، و نتایج هر کدام را توضیح دهید.

برای طراحی این طبقه بند ابتدا پس از load کردن تصاویر، میانگین مقدار RGB پیکسل‌های تصاویر را برای هرکدام محاسبه کرده و در یک جدول ذخیره میکنیم. در مرحله بعد به کمک ۲ طبقه بند(یکی کتابخانه و یکی توسعه داده شده) تصاویر را طبقه میکنیم. طبقه بند اول از Multinomial bayes استفاده میکنیم(از این طبقه بند برای داده‌هایی با فیچرهای مجزا استفاده میشود). نتایج تمامی آزمایشات درون فایل نوت‌بوک قابل مشاهده هستند.

```
Accuracy of model is:  0.8979591836734694
Percision of model is:  0.8888888888888888
Recall of model is:  0.9230769230769231
Confusion matrix of model is:
 [[20  3]
 [ 2 24]]
```

برای طبقه بند دوم، مقادیر رنگ‌های قرمز و آبی را مدنظر قرار میدهیم و هر رنگی بیشتر بود، تیم را به آن رنگ نسبت میدهیم.

```
Accuracy of model is:  0.8442622950819673
Percision of model is:  0.9787234042553191
Recall of model is:  0.71875
Confusion matrix of model is:
 [[57  1]
 [18 46]]
```

سوال ششم

در این سوال قصد داریم دیتاست Iris را بررسی کنیم. در این دیتاست سه ویژگی برای هر کلاس داده شده است و در کل داده ها به سه کلاس تقسیم می شوند.) 20 نمره)

الف)ابتدا توضیح مختصری راجع به طبقه بندهای naïve bayes و optimal bayes داده و آن ها را با فرض گاوسی بودن داده های ورودی پیادهسازی کنید. دقت کنید که الگوریتم شما نباید هیچ پیش فرضی بر از دادههای ورودی داشته باشد و باید برای هر دادهای کار کند. (در این قسمت مجاز به استفاده از کتابخانه های آماده نیستید و باید الگوریتم را خودتان پیاده سازی کنید.)

ب)الگوریتم های پیادهسازی شده را برای مجموعه داده Iris تست کنید و نتایج به دست آمده (Confusion matrix ، دقت طبقه بند) را برای هر روش گزارش کنید و در مورد نتایج بحث کنید.

ج) در این قسمت با استفاده از پکیج های آماده پایتون فقط الگوریتم naïve bayes را با فرض گاوسی بودن اجرا کنید و نتیجه را گزارش کنید.

قضیه بیز روشی را ارائه می دهد که ما می توانیم با توجه به دانش قبلی خود، احتمال یک قطعه داده متعلق به یک کلاس معین را محاسبه کنیم. قضیه بیز به صورت زیر بیان می شود:

$$P(C|D) = (P(D|C) * P(C)) / P(D)$$

Naive Bayes یک الگوریتم طبقه بندی برای مسائل طبقه بندی باینری (دو کلاسه) و چند کلاسه است. به آن بیز ساده نیز میگویند زیرا محاسبات احتمالات برای هر کلاس ساده شده است تا محاسبات آنها قابل انجام باشد.

به جای تلاش برای محاسبه احتمالات هر مقدار مشخصه، فرض می شود که با توجه به مقدار کلاس، آنها به صورت شرطی مستقل هستند.

این یک فرض بسیار قوی است که در داده های واقعی بسیار بعید است(یعنی ویژگی ها با هم تعامل نداشته باشند). با این وجود، این رویکرد به طرز شگفتآوری روی دادههایی که برای آنها این فرض صادق نیست، خوب عمل می‌کند.

طبقه‌بند بهینه Bayes یک مدل احتمالی است که با توجه به مجموعه داده‌های آموزشی، محتمل‌ترین پیش‌بینی را برای یک مثال جدید انجام می‌دهد.این مدل همچنین به عنوان یادگیرنده بهینه بیز، طبقه بندی کننده بیز، مرز تصمیم گیری بهینه بیز یا تابع تفکیک بهینه بیز نیز شناخته می شود.

معادله زیر نحوه محاسبه احتمال شرطی برای یک نمونه جدید (vi) را با توجه به داده های آموزشی (D) و با توجه به فضایی از فرضیه ها (H) نشان می دهد.

$$P(vj \mid D) = \text{sum } \{h \text{ in } H\} P(vj \mid hi) * P(hi \mid D)$$

در این فرمول vj یک نمونه جدید برای طبقه‌بندی است، H مجموعه‌ای از فرضیه‌ها برای طبقه‌بندی نمونه است، hi یک فرضیه است، P(vj | hi) احتمال پسین برای vi با توجه به فرضیه داده شده hi، و P(hi | D) است. احتمال پیشین فرضیه hi با توجه به داده های D است.انتخاب نتیجه به کمک maximum likelihood انجام میگیرد.

max sum {h in H} P(vj | hi) * P(hi | D)

هیچ روش طبقه بندی دیگری با استفاده از فضای فرضی مشابه و دانش قبلی نمی تواند به طور متوسط از این روش بهتر عمل کند.

این بدان معناست که هر الگوریتم دیگری که بر روی همان داده‌ها، مجموعه فرضیه‌های مشابه و احتمالات پیشین مشابه عمل می‌کند، به طور متوسط نمی‌تواند از این رویکرد بهتر عمل کند. از این رو نام این طبقه بند،" بهینه" است.

اگرچه این طبقه‌بند پیش‌بینی‌های بهینه انجام می‌دهد، اما با توجه به عدم قطعیت در داده‌های آموزشی و پوشش ناقص حوزه مسئله و فضای فرضیه، کامل نیست. به این ترتیب، مدل خطا خواهد داشت. این خطاها اغلب به عنوان خطاهای Bayes شناخته می شوند.

برای حل این مساله ۳ پیاده سازی داخل نوت‌بوک انجام شده. اولی طبقه بند naiiv bayes توسعه داده شده، دومی طبقه بند naiiv bayes به کمک کتابخانه sklearn و مورد سوم هم طبقه بند optimal classifier که با توجه به فرض گاوسی بودن از فرمول زیر پیروی میکند:

$$f_x(x_1, \dots, x_n) = \frac{exp(-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu))}{\sqrt{(2\pi)^n|\Sigma|}}$$

نتایج حاصل از آزمایش این ۳ روش نشان داد که الگوریتم های optimal و naiive برای این مساله پاسخ تقریبا یکسانی دارند و مقدارکمی از نتایج کتابخانه بهتر عمل کردند.