فردین عباسی 810199456

به نام خدا

تمرین سری اول درس یادگیری ماشین

دانشکده مهندسی برق و کامپیوتر دانشگاه تهران

زمستان 1401

# تمرین سری 1

سوال 1

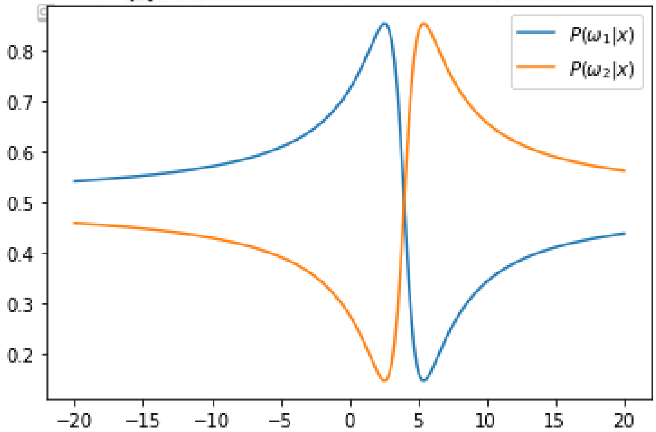
آ)

$$P(x|\omega_i) = \frac{1}{\pi b}\,\frac{1}{1+\left(\frac{x-a_i}{b}\right)^2} \quad i=1,2 \quad a_2>a_1$$

آ) $P(\omega_1)=P(\omega_2)$ $\quad P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}$ $\quad$ if $P(\omega_1)=P(\omega_2)$ then $\frac{P(\omega_1|x)}{P(x|\omega_1)} = \frac{P(\omega_2|x)}{P(x|\omega_2)}$

if $P(\omega_1|x)=P(\omega_2|x)$ then $P(x|\omega_1)=P(x|\omega_2)$

$$\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2} = \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} \rightarrow a_1=a_2 \;\times\; \text{or}\; x-a_1=a_2-x \rightarrow x=\frac{a_1+a_2}{2}$$




ب)

ب) $P(error) = \int_{\omega_1} P(x|\omega_2)P(\omega_2) + \int_{\omega_2} P(x|\omega_1)P(\omega_1) = \frac{1}{\pi b}\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{0.5}{1+\left(\frac{x-a_2}{b}\right)^2}dx$

$+ \frac{1}{\pi b}\int_{\frac{a_1+a_2}{2}}^{\infty} \frac{0.5}{1+\left(\frac{x-a_1}{b}\right)^2}dx = \frac{1}{\pi b}\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{dx}{1+\left(\frac{x-a_2}{b}\right)} \xrightarrow{y=\frac{x-a_2}{b}} = \frac{1}{\pi}\int_{-\infty}^{\frac{a_1-a_2}{2b}} \frac{dy}{1+y^2}$

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

پ)

پ) $P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right| \quad \therefore -\frac{\pi}{2} < \tan^{-1}\left|\frac{a_2-a_1}{2b}\right| < \frac{\pi}{2} \rightarrow 0 < P(error) < \frac{1}{2}$

$\max P(error) = \frac{1}{2} \Rightarrow \tan^{-1}\left|\frac{a_2-a_1}{2b}\right| = \frac{\pi}{2}\; \begin{cases} a_2=a_1 \\ b\rightarrow\pm\infty \end{cases}$

ت)

ت) $P(w_1|x) \underset{2}{\overset{1}{\gtrless}} P(w_2|x) \xrightarrow[\text{rule}]{\text{bayes}} \dfrac{P(x|w_1)}{P(x|w_2)} \underset{2}{\overset{1}{\gtrless}} \dfrac{P(w_2)}{P(w_1)}$ Decision boundary

$\rightarrow \dfrac{1+\left(\frac{x-a_2}{b}\right)^2}{1+\left(\frac{x-a_1}{b}\right)^2} \underset{2}{\overset{1}{\gtrless}} 1 \rightarrow |x-a_2| \gtrless |x-a_1| \rightarrow x = \dfrac{a_1+a_2}{2}$

مرز تصمیم جایی است که احتمال پسین هر دو کلاس برابر باشد.

همانطور که در قسمت ب نشان دادیم در مرز تصمیم میزان احتمال خطا حداقل و برابر است با:

$$P(error) = 0.5 - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|$$

ث)

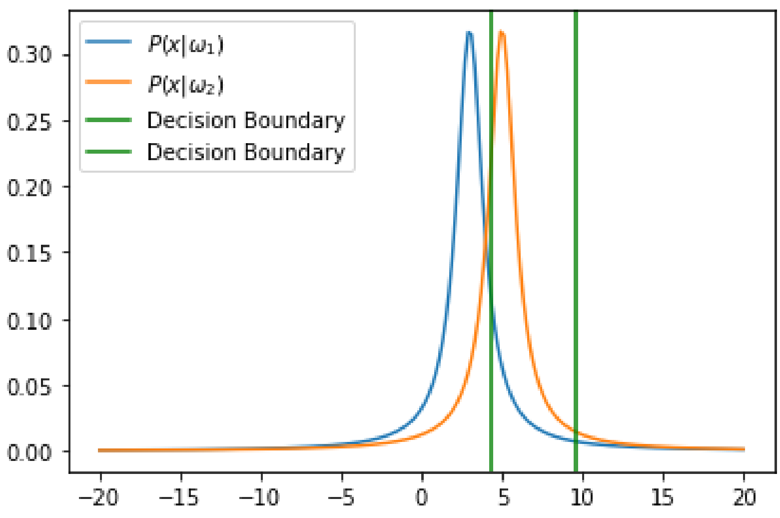
$$Decision\ Boundary: \lambda_{12}P(w_2|x) = \lambda_{21}P(w_1|x) \rightarrow \frac{P(x|w_2)}{P(x|w_1)} = \frac{\lambda_{21}}{\lambda_{12}} = 2 \rightarrow \frac{1+\left(\frac{x-a_1}{b}\right)^2}{1+\left(\frac{x-a_2}{b}\right)^2} = 2$$

$$x = 2a_2 - a_1 \pm \sqrt{2a_2^2 + 2a_1^2 - 4a_1a_2 - b^2} = 9.64\&4.35$$

در بازه $2a_2 - a_1 - \sqrt{2a_2^2 + 2a_1^2 - 4a_1a_2 - b^2} < x < 2a_2 - a_1 + \sqrt{2a_2^2 + 2a_1^2 - 4a_1a_2 - b^2}$ متعلق به کلاس 2 و خارج از آن کلاس 1 است

$$P(eror) = \int_{w1} P(x|w2)P(w2) + \int_{w2} P(x|w1)P(w1) = \frac{1}{\pi b}\int_{-\infty}^{2a_2-a_1-\sqrt{2a_2^2+2a_1^2-4a_1a_2-b^2}} \frac{0.5}{1+\left(\frac{x-a_2}{b}\right)^2}dx$$

$$+\frac{1}{\pi b}\int_{2a_2-a_1-\sqrt{2a_2^2+2a_1^2-4a_1a_2-b^2}}^{2a_2-a_1+\sqrt{2a_2^2+2a_1^2-4a_1a_2-b^2}} \frac{0.5}{1+\left(\frac{x-a_1}{b}\right)^2}dx + \frac{1}{\pi b}\int_{2a_2-a_1+\sqrt{2a_2^2+2a_1^2-4a_1a_2-b^2}}^{\infty} \frac{0.5}{1+\left(\frac{x-a_2}{b}\right)^2}dx = 0.26$$



## سوال 2

$P(w_1)=P(w_2)$  Decision Boundary: $P(w_1|x)=P(w_2|x) \to P(x|w_1)=P(x|w_2)$

$\frac{x}{\sigma_1^2}\exp\left(\frac{-x^2}{2\sigma_1^2}\right)=\frac{x}{\sigma_2^2}\exp\left(\frac{-x^2}{2\sigma_2^2}\right)\xrightarrow{\ln} -2\ln\sigma_1-\frac{x}{2\sigma_1^2}=-2\ln\sigma_2-\frac{x}{2\sigma_2^2}$

$\to x=\frac{4\sigma_1^2\sigma_2^2\ln\sigma_1/\sigma_2}{\sigma_1^2-\sigma_2^2}$



## سوال 3

آ)

چون درباره توزیع پیشین کلاس ها اطلاعاتی نداریم احتمال هر یک را برابر تعداد نقاط شان به کل در نظر می گیریم:

$$P(w_1)=\frac{10}{19}\,,P(w_2)=\frac{9}{19} \qquad g_i(x)=X^TW_iX+w_i^TX+w_{i0} \qquad X=\begin{bmatrix}x\\y\end{bmatrix}$$

$$Wi=-0.5\times\Sigma_i^{-1} \qquad w_i=\Sigma_i^{-1}\mu_i \qquad w_{i0}=-0.5\times\mu_i^T\Sigma_i^{-1}\mu_i-\frac{ln|\Sigma_i|}{2}+\ln\left(P(w_i)\right)$$

$$\mu_1=\begin{bmatrix}-0.15\\-0.15\end{bmatrix},\mu_2=\begin{bmatrix}1.33\\1.61\end{bmatrix} \qquad \Sigma_1=\begin{bmatrix}1.65 & 0.0025\\0.0025 & 0.503\end{bmatrix},\Sigma_2=\begin{bmatrix}0.55 & 0.1852\\0.1852 & 0.98\end{bmatrix}$$

$$W_1=\begin{bmatrix}-0.303 & 0.0015\\0.0015 & -1\end{bmatrix} \qquad w_1=\begin{bmatrix}-0.091\\-0.298\end{bmatrix} \qquad w_{10}=-0.578$$

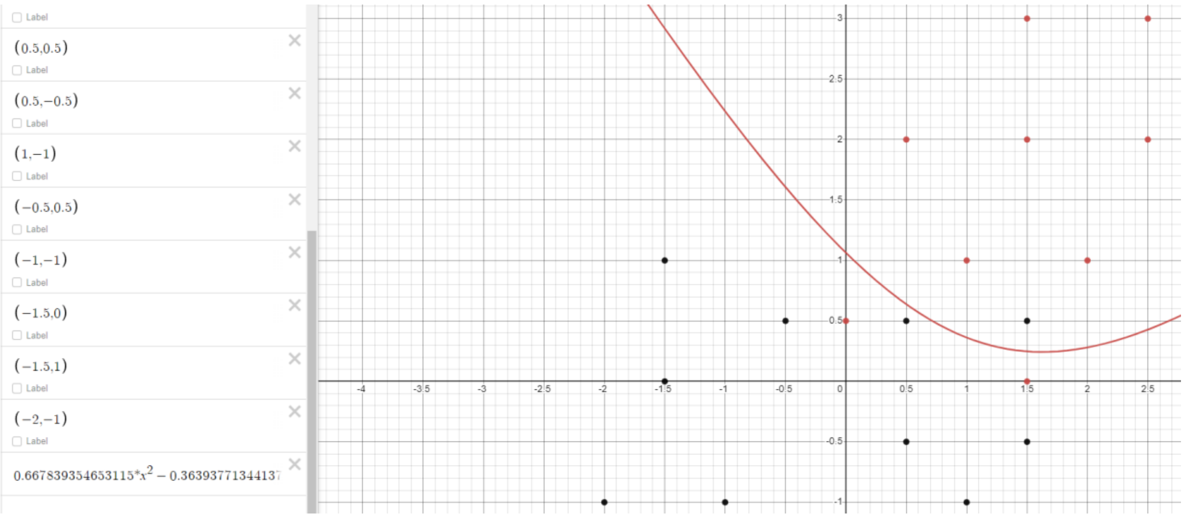
$$g_1(x)=-0.303x^2+0.003xy-0.09x-y^2-0.3y-0.578$$

$$W_2=\begin{bmatrix}-0.97 & 0.18\\0.18 & -0.545\end{bmatrix} \qquad w_2=\begin{bmatrix}2\\1.26\end{bmatrix} \qquad w_{20}=-2.75$$

$$g_2(x)=-0.97x^2+0.367xy+2x-0.545y^2+1.26y-2.75$$

$Decision\ Boundary$: $g_1(x)=g_2(x)$

$$0.67x^2-0.363xy-2.08x-0.45y^2-1.565y+2.17=0$$



ب)

$$\mu_1 = \frac{1}{10}\left(\begin{bmatrix}-2\\-1\end{bmatrix}+\begin{bmatrix}-1.5\\0\end{bmatrix}+\begin{bmatrix}-1.5\\1\end{bmatrix}+\begin{bmatrix}-1\\-1\end{bmatrix}+\begin{bmatrix}-0.5\\0.5\end{bmatrix}+\begin{bmatrix}0.5\\0.5\end{bmatrix}+\begin{bmatrix}0.5\\-0.5\end{bmatrix}+\begin{bmatrix}1\\-1\end{bmatrix}+\begin{bmatrix}1.5\\0.5\end{bmatrix}+\begin{bmatrix}1.5\\-0.5\end{bmatrix}\right) = \begin{bmatrix}0.15\\-0.15\end{bmatrix}$$

$$\mu_2 = \frac{1}{9}\left(\begin{bmatrix}0\\0.5\end{bmatrix}+\begin{bmatrix}0.5\\2\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}1.5\\0\end{bmatrix}+\begin{bmatrix}1.5\\2\end{bmatrix}+\begin{bmatrix}2\\1\end{bmatrix}+\begin{bmatrix}2\\3\end{bmatrix}+\begin{bmatrix}2.5\\2\end{bmatrix}\right) = \begin{bmatrix}1.33\\1.61\end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix}Var\,x & cov\,x,y\\ cov\,y,x & Var\,y\end{bmatrix}, \quad Var\,x = \frac{1}{10}\sum_{i=1}^{10}(x-\mu_x)^2 = 1.4425$$

$$Var\,y = \frac{1}{10}\sum_{i=1}^{10}(y-\mu_y)^2 = 0.5025$$

$$Cov_{x,y} = \frac{\sum_{i=1}^{10}(x-\mu_x)(y-\mu_y)}{N} = 0.0025 \qquad \Sigma_1 = \begin{bmatrix}1.4425 & 0.0025\\ 0.0025 & 0.5025\end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix}Var\,x & Cov\,x,y\\ cov\,y,x & Var\,y\end{bmatrix} \quad Cov\,x,y = \frac{\sum_{i=1}^{9}(x-\mu_x)(y-\mu_y)}{N} = 0.1852 \quad Var\,x = 0.55$$

$$Var\,y = 0.98 \qquad \Sigma_2 = \begin{bmatrix}0.55 & 0.1852\\ 0.1852 & 0.98\end{bmatrix}$$

پ)

$$W_1 = \begin{bmatrix}-0.303 & 0.0015\\ 0.0015 & -1\end{bmatrix} \qquad w_1 = \begin{bmatrix}-0.091\\-0.298\end{bmatrix} \qquad w_{10} = -0.63$$

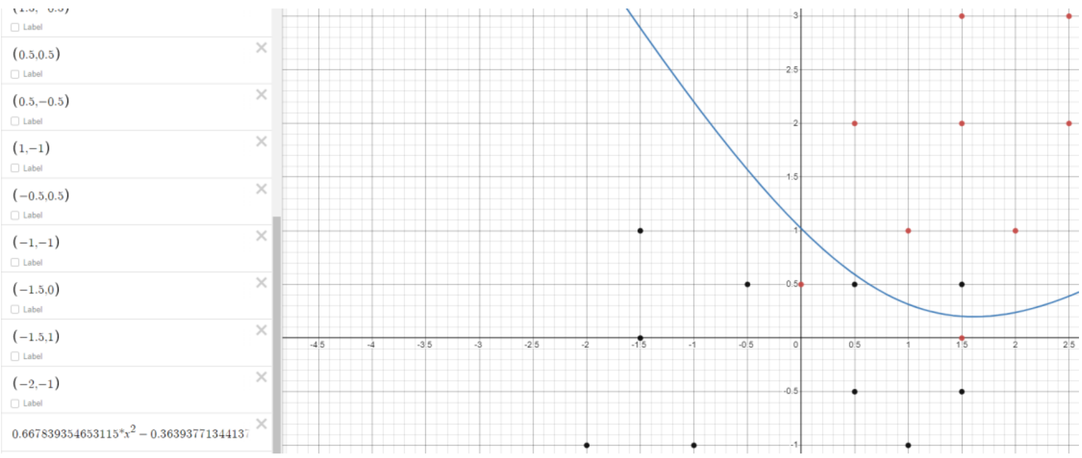
$$g_1(x) = -0.303x^2 + 0.003xy - 0.09x - y^2 - 0.3y - 0.63$$

$$W_2 = \begin{bmatrix}-0.97 & 0.18\\ 0.18 & -0.545\end{bmatrix} \qquad w_2 = \begin{bmatrix}2\\1.26\end{bmatrix} \qquad w_{20} = -2.7$$

$$g_2(x) = -0.97x^2 + 0.367xy + 2x - 0.545y^2 + 1.26y - 2.7$$

$$P(w_1) = P(w_2) \rightarrow g_1(x) = g_2(x)$$

$$Decision\ Boundary: 0.66x^2 - 0.36xy - 2.08x - 0.45y^2 - 1.57y + 2.06 = 0$$




خطای آموزش

تجربی 3/19=15%

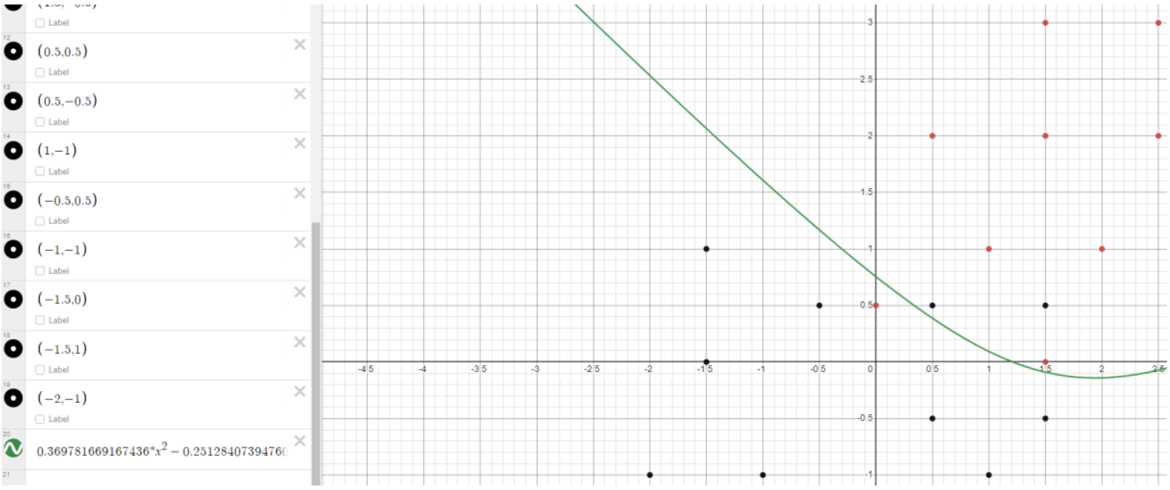
(ت

$R(i|x) = \sum \lambda(i|j) P(j|x)$ $\qquad$ $Decision\ Boundary: \lambda(w_1|w_2)P(w_2|x) = \lambda(w_1|w_2)P(w_1|x)$

$$\Longrightarrow \frac{P(x|w_2)}{P(x|w_1)} = \frac{\lambda(w_2|w_1)}{\lambda(w_1|w_2)} \times \frac{P(w_1)}{P(w_2)} \xrightarrow{P(w_1)=P(w_2)} \frac{g_1(x)}{g_2(x)} = ln2$$

$$Decision\ Boundary: 0.37x^2 - 0.25xy - 1.47x - 0.616y^2 - 1.176y + 1.238 = 0$$

با توجه به اینکه هزینه کلاس 1 را 2 گفتن بیش تر از کلاس 2 را 1 گفتن است مشاهده می کنیم مرز تصمیم به بالاتر رفته تا هیچ کلاس 1 ای اشتباه کلاس بندی نشود.




(ث

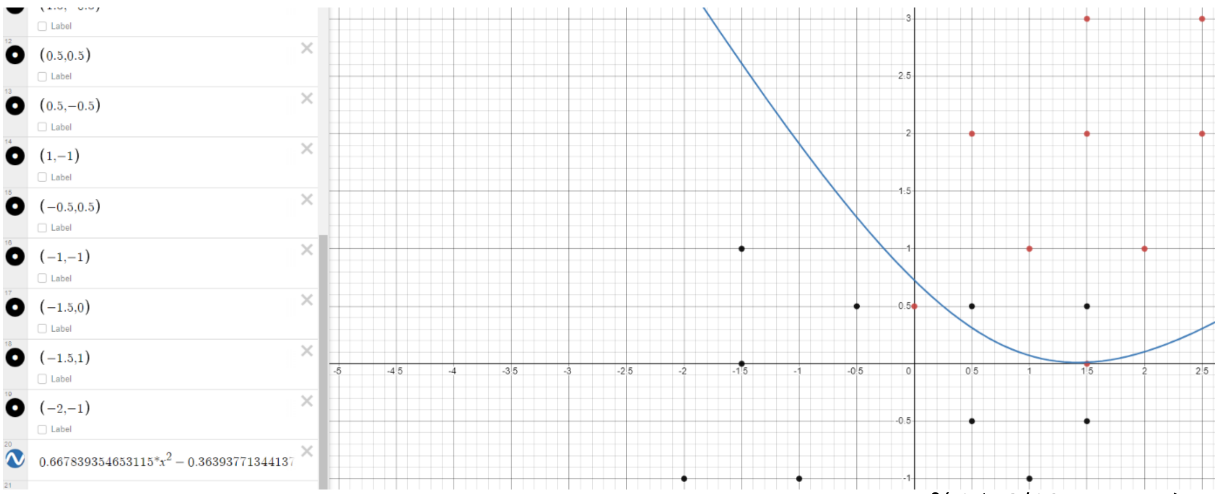
$$W_1 = \begin{bmatrix} -0.303 & 0.0015 \\ 0.0015 & -1 \end{bmatrix} \qquad w_1 = \begin{bmatrix} -0.091 \\ -0.298 \end{bmatrix} \qquad w_{10} = -1.03$$

$$g_1(x) = -0.303x^2 + 0.003xy - 0.09x - y^2 - 0.3y - 1.03$$

$$W_2 = \begin{bmatrix} -0.97 & 0.18 \\ 0.18 & -0.545 \end{bmatrix} \qquad w_2 = \begin{bmatrix} 2 \\ 1.26 \end{bmatrix} \qquad w_{20} = -2.41$$

$$g_2(x) = -0.97x^2 + 0.367xy + 2x - 0.545y^2 + 1.26y - 2.41$$

$$g_1(x) = g_2(x) \ \ Decision\ Boundary: 0.67x^2 - 0.36xy - 2.08x - 0.45y^2 - 1.565y + 1.37 = 0$$




خطای آموزش تجربی= 3/19=15%

سوال 4

الف) $P(D|\lambda) = P(x_1, x_2, \dots, x_n|\lambda) \overset{i.i.d}{=} \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \xrightarrow{\ln} \sum_{i=1}^{n} x_i \ln\lambda - \lambda - \ln x_i! = \log(\text{Likelihood})$

$$\frac{\partial L(\lambda)}{\partial \lambda} = \sum_{i=1}^{n} \frac{x_i}{\lambda} - 1 = 0 \rightarrow \sum_{i=1}^{n} \frac{x_i}{\lambda} = n \rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \lambda_{MLE}$$

ب)

$$P(\lambda|D) = \frac{P(D|\lambda)\,P(\lambda)}{\int P(D|\lambda)P(\lambda)d\lambda} = c\,\mathrm{Gamma}(\lambda|\alpha,\beta) \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= c'\lambda^{\alpha-1} e^{-\beta\lambda} \times \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-\lambda n}}{\prod_{i=1}^{n} x_i!} = \frac{c'\lambda^{\alpha-1+\sum_{i=1}^{n} x_i} \times e^{-\lambda(\beta+n)}}{\prod_{i=1}^{n} x_i!} = \mathrm{Gamma}(\lambda|\alpha+n\bar{x}_i, \beta+n)$$

$c''$

where $\bar{x}_i = \frac{\sum_{i=1}^{n} x_i}{n}$ & $c'' = \int_{-\infty}^{\infty} \mathrm{Gamma}(\lambda|\alpha+n\bar{x}_i, \beta+n)d\lambda$

پ) بله زیرا توزیع Prior، $P(\lambda)$ و Posterior، $P(\lambda|D)$ هر دو از گاما پیروی کنند هم جنس است.

ت) $\hat{\lambda}_{MAP} = \arg\max_{\lambda} P(D|\lambda)P(\lambda)$ where $\lambda_{MAP} = \frac{\alpha+n\bar{x}_i-1}{\beta+n}$

اگر $n\rightarrow\infty$: $\hat{\lambda}_{MAP} = \bar{x}_i = \frac{\sum_{i=1}^{n} x_i}{n} = \lambda_{MLE}$ اگر $n\rightarrow\infty$، $\lambda$ حاصل از تخمین MAP و MLE برابر است

و این به این دلیل است که $\lambda$ حاصل از MLE fixed point است ولی $\lambda$ حاصل از MAP یک توزیع است که هنگامی که $n\rightarrow\infty$ این توزیع یک تابع ضربه به جای fixed point است.

ج) معمولاً هنگامی که dataset کوچک است تخمین گر MAP به دلیل در نظر گرفتن Prior گزینه بهتری است اما هنگامی که dataset به حد کافی بزرگ است، استفاده از MLE بهتر است زیرا معمولاً یافتن توزیع Prior کار پیچیده‌ای است

## سوال 5

آ)

در bayes classifier به ازای L تا فیچر به تعداد $N^L$ دیتاپوینت نیاز داریم تا مدل دقیق کار کند ولی اغلب جمع آوری دیتا فرآیندی پرهزینه و زمان بر است پس همواره به دنبال راهکاری برای کاهش آن هستیم.

در naïve bayes فیچر ها را نسبت به هم مستقل می گیریم واین کار سبب میشود به ازای L تا فیچر تنها به NL تا دیتاپوینت نیاز داشته باشیم.این دیدگاه سبب می شود علی رغم افت اندک دقت مدل، با تعداد دیتا به مراتب کمتر دقت نسبتا خوبی از مدل بگیریم و اغلب هم پرکاربرد است.

در پیش پردازش داده های بدون مقدار را با میانگین جایگزین کرده سپس ضمن shuffle کردن داده های test & train را جدا می کنیم.

ب)

توضیح روند کد:

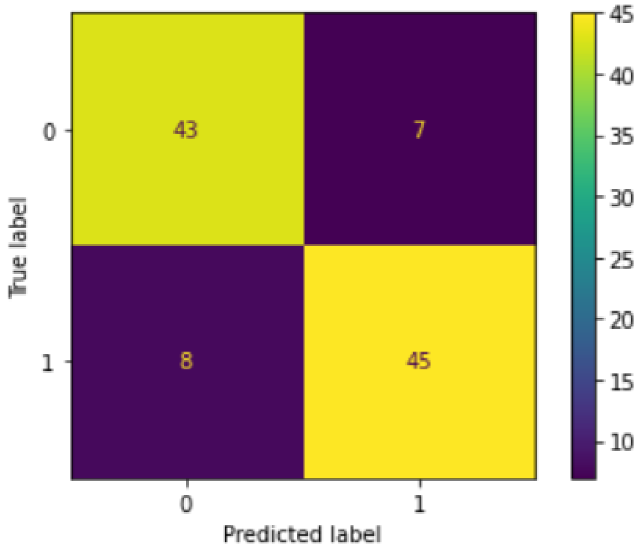
برای هر کلاس داده های train با محاسبه میانگین و واریانس، توزیع گوسی هر کلاس را تشکیل می دهیم. سپس به ازای هر داده test میزان توزیع گوسی هر کلاس را محاسبه و بیشترین را به عنوان لیبل حدس زده شده در نظر می گیریم.

در 1 vs. ALL یک کلاس را نگه داشته و بقیه کلاس ها را یک کلاس در نظر میگیریم. توجه: در سوال 5 همواره فرض بر این شد که کلاس هدف others است و حدس درست آن true positive و حدس درست کلاس دیگر true negative است.
واضح است با توجه به داشتن ماتریس آشفتگی یافتن مقادیر دقت و... در صورت هدف بودن کلاس دیگری قابل محاسبه است

### Adelie vs. Others

کلاس Adelie عدد باینری 0 و Other عدد باینری 1 است

```
Accuracy= 85.43689320388835%
Recall= 84.90566037735849%
Percision= 86.53846153846153%
```

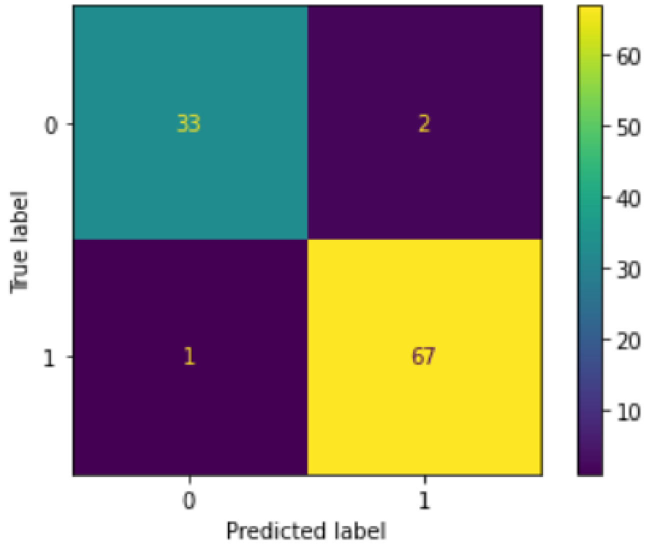



در مجموع طبقه بند با accuracy= 85%, recall=85%,precision= 86.5% عملکرد خوبی دارد.

## Gentoo vs. Others

کلاس Gentooعدد باینری 0 و Others عدد باینری 1 است.

```
Accuracy= 97.0873786407767%
Recall= 98.52941176470588%
Percision= 97.10144927536231%
```




در مجموع طبقه بند با accuracy= 97%, recall=99%,precision= 97% عملکرد عالی دارد.

## Chinstrap vs. Others

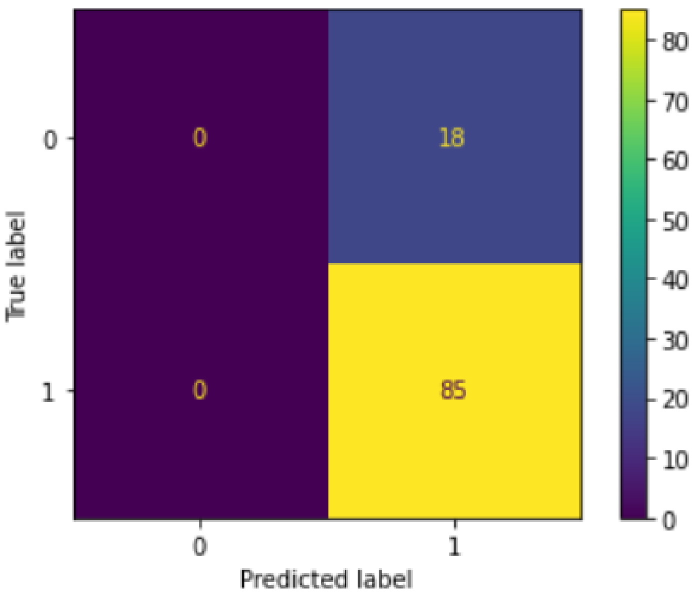
کلاس Chinstrapعدد باینری 0 و Others عدد باینری 1 است.

اینجا طبقه بند تمامی داده ها را others لیبل می زند.

علت این خطا به دلیل اکیدا کوچک بودن احتمال پرایور chinstrap در مقابل بقیه کلاس ها است.

به همین دلیل باز هم از مدل دقت خوب 82 درصدی دریافت می کنیم.

```
Accuracy= 82.52427184466019%
Recall= 100.0%
Percision= 82.52427184466019%
```

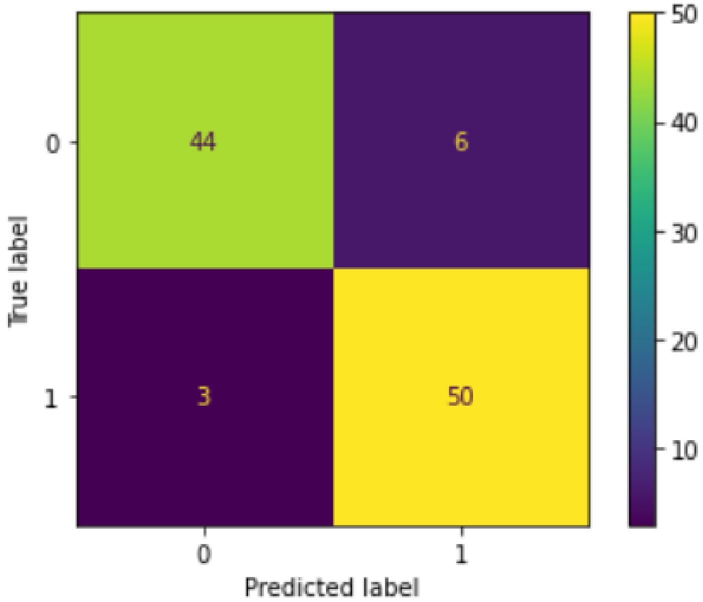


پ)

### Adelie vs. Others

کلاس Adelie عدد باینری 0 و Other عدد باینری 1 است

```
Accuracy= 91.2621359223301%
Recall= 94.33962264150944%
Percision= 89.28571428571429%
```



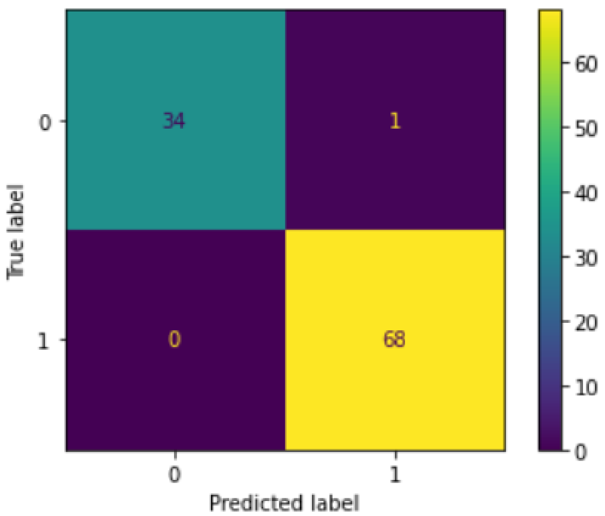

با استفاده از کتابخانه SKLEARN

عملکرد طبقه بند به طور مشهود پیشرفت کرده است.

### Gentoo vs. Others

کلاس Gentoo عدد باینری 0 و Other عدد باینری 1 است

```
Accuracy= 99.02912621359224%
Recall= 100.0%
Percision= 98.55072463768116%
```



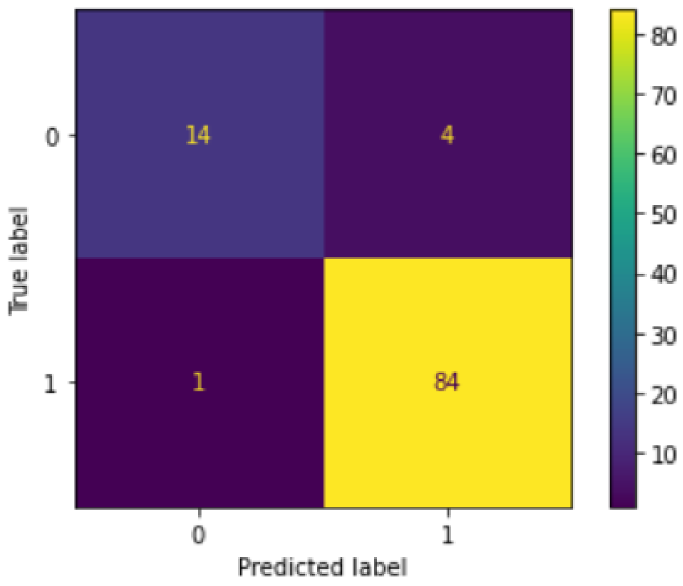

با استفاده از کتابخانه SKLEARN

عملکرد طبقه بند به طور مشهود پیشرفت کرده است.

**Chinstrap vs. Others**

کلاس Gentoo عدد باینری 0 و Other عدد باینری 1 است

```
Accuracy= 95.14563106796116%
Recall= 98.82352941176471%
Percision= 95.45454545454545%
```



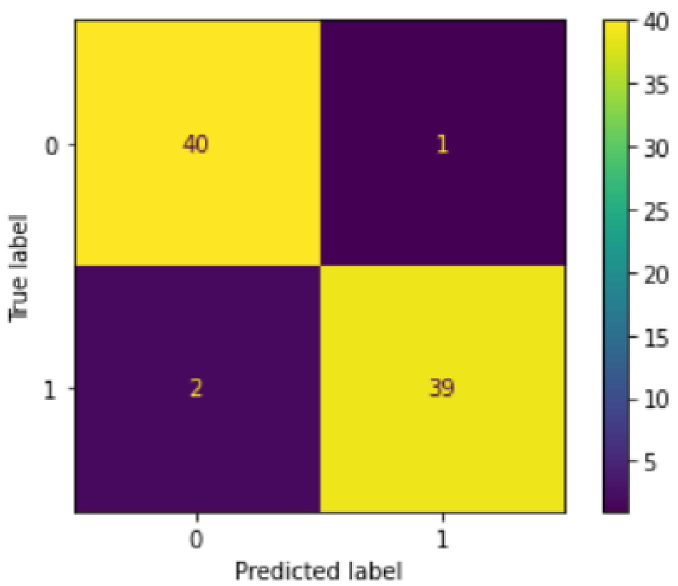

در اینجا مشکل طبقه بند دستی برطرف شده است و مدل عملکرد خوبی به نمایش گذاشته است.

## سوال 6

در دیتاست صورت تمرین اسم فایل تصاویر جنگل ها از الگوی 'image/s'+str(i)+'.jpg' پیروی می کند جز دو تصویر آخر که چند شماره جلو می افتد، نام این تصاویر قبل از اجرای کد مرتب شد.

با فرض کلاس هدف دریا،ماتریس آشفتگی و مقادیر دقت و... به شرح زیر است:

```
Accuracy= 96.34146341463415%
Recall= 95.1219512195122%
Percision= 0.975%
```




یک تصویر دریا به اشتباه جنگل حدس زده شد و دو تصویر جنگل به اشتباه دریا حدس زده شد.

این تصویر به اشتباه جنگل حدس زده شد که به علت وجود انبوه رنگ سبز در پلات تصویر می باشد

دو تصویر زیر نیز به دلیل انبوه رنگ آبی در پلات آن به اشتباه دریا حدس زده شد