ECE M146 - HW 6

May 22, 2020

John Rapp Farnes | 405461225

1

1.1 (a)

We have

$$\Sigma = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \implies$$
$$\Sigma^{-1} = \frac{1}{\det \Sigma} \begin{bmatrix} \alpha_4 & -\alpha_2 \\ -\alpha_3 & \alpha_1 \end{bmatrix}$$
$$\det \Sigma = \alpha_1 \alpha_4 - \alpha_2 \alpha_3$$

Writing out the multivariate form of the pdf with two random gaussians X and Y:

$$f_Z(x,y) = \underbrace{\frac{1}{\sqrt{(2\pi)^2 \det \Sigma}}}_{C} \exp\Big\{ -\frac{1}{2}\Big(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \Big)^T \frac{1}{\det \Sigma} \begin{bmatrix} \alpha_4 & -\alpha_2 \\ -\alpha_3 & \alpha_1 \end{bmatrix} \Big(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \Big) \Big\}$$

Starting with the constant C, we can write

$$\begin{aligned} C &= 2\pi \sqrt{\alpha_1 \alpha_4 - \alpha_2 \alpha_3} \\ &= 2\pi \sqrt{\alpha_1 \alpha_4 [1 - \frac{\alpha_2^2}{\alpha_1 \alpha_4}]} \\ &= 2\pi \sqrt{\alpha_1 \alpha_4} \sqrt{1 - \frac{\alpha_2 \alpha_3}{[\sqrt{\alpha_1 \alpha_4}]^2}} \end{aligned}$$

Comparing to the expression of the constant of the two jointly Gaussan pdf C', we get:

$$C = C' \Leftrightarrow$$
$$2\pi \sqrt{\alpha_1 \alpha_4} \sqrt{1 - \frac{\alpha_2 \alpha_3}{[\sqrt{\alpha_1 \alpha_4}]^2}} = 2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho_{X,Y} \rho_{X,Y}}$$

As such, we get equality if we let $\alpha_1 = \sigma_1$, $\alpha_2 = \sigma_2$ and $\alpha_3 = \alpha_4 = \sigma_1 \sigma_2 \rho_{X,Y}$

Continuing with the exp expression:

$$
\begin{aligned}
f_Z(x,y) &= \frac{1}{C}\exp\Big\{-\frac{1}{2}\begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}^T \frac{1}{\sigma_1^2\sigma_2^2(1-\rho_{X,Y}^2)} \begin{bmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho_{X,Y} \\ -\sigma_1\sigma_2\rho_{X,Y} & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}\Big\} \\
&= \frac{1}{C}\exp\Big\{-\frac{1}{2(1-\rho_{X,Y}^2)}\frac{1}{\sigma_1^2\sigma_2^2} \underbrace{\begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}^T \begin{bmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho_{X,Y} \\ -\sigma_1\sigma_2\rho_{X,Y} & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}}_{=Q}\Big\}
\end{aligned}
$$

Expanding the quadratic form Q:

$$
\begin{aligned}
Q &= \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}^T \begin{bmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho_{X,Y} \\ -\sigma_1\sigma_2\rho_{X,Y} & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix} \\
&= \begin{bmatrix} x-\mu_1 & y-\mu_2 \end{bmatrix} \begin{bmatrix} \sigma_2^2 x - \sigma_2^2\mu_1 - \sigma_1\sigma_2\rho_{X,Y}y + \sigma_1\sigma_2\rho_{X,Y}\mu_2 \\ -\sigma_1\sigma_2\rho_{X,Y}x + \sigma_1\sigma_2\rho_{X,Y}\mu_1 + \sigma_1^2 y - \sigma_1^2\mu_2 \end{bmatrix} \\
&= \big[(x-\mu_1)(\sigma_2^2 x - \sigma_2^2\mu_1 - \sigma_1\sigma_2\rho_{X,Y}y + \sigma_1\sigma_2\rho_{X,Y}\mu_2) + (y-\mu_2)(-\sigma_1\sigma_2\rho_{X,Y}x + \sigma_1\sigma_2\rho_{X,Y}\mu_1 + \sigma_1^2 y - \sigma_1^2\mu_2)\big] \\
&= (x-\mu_1)\big[\sigma_2^2(x-\mu_1) - \rho_{X,Y}\sigma_1\sigma_2(y-\mu_2)\big] + (y-\mu_2)\big[-\sigma_1\sigma_2\rho_{X,Y}(x-\mu_1) + \sigma_1^2(y-\mu_2)\big] \\
&= \sigma_2^2(x-\mu_1)^2 - 2\sigma_1\sigma_2\rho_{X,Y}(x-\mu_1)(y-\mu_2) + \sigma_1^2(y-\mu_2)^2
\end{aligned}
$$

Putting it together, we get:

$$
\begin{aligned}
f_Z(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}\exp\Big\{-\frac{1}{2(1-\rho_{X,Y}^2)}\Big[\frac{\sigma_2^2(x-\mu_1)^2 - 2\sigma_1\sigma_2\rho_{X,Y}(x-\mu_1)(y-\mu_2) + \sigma_1^2(y-\mu_2)^2}{\sigma_1^2\sigma_2^2}\Big]\Big\} \\
&= \frac{\exp\Big\{\frac{-1}{2(1-\rho_{X,Y}^2)}\Big[\big(\frac{x-\mu_1}{\sigma_1}\big)^2 - 2\rho_{X,Y}\big(\frac{x-\mu_1}{\sigma_1}\big)\big(\frac{y-\mu_2}{\sigma_2}\big) + \big(\frac{y-\mu_2}{\sigma_2}\big)^2\Big]\Big\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}
\end{aligned}
$$

Finally, letting $\mu_1 = m_1$ and $\mu_2 = m_2$, we get $f_Z = f_{X,Y}$

$$
\therefore
\begin{cases}
\mu = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{X,Y} \\ \sigma_1\sigma_2\rho_{X,Y} & \sigma_2^2 \end{bmatrix} \\
\Sigma^{-1} = \begin{bmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho_{X,Y} \\ -\sigma_1\sigma_2\rho_{X,Y} & \sigma_1^2 \end{bmatrix}
\end{cases}
$$

1.2 (b)

$\rho_{X,Y} = 0 \implies \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$

Further, we have

$$
\begin{aligned}
f_{X,Y}(x,y) &= \frac{\exp\left\{ \frac{-1}{2}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 + \left(\frac{y-m_2}{\sigma_2}\right)^2 \right] \right\}}{2\pi\sigma_1\sigma_2} \\
&= \frac{\exp\left\{ -\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2 \right\}}{\sigma_1\sqrt{2\pi}} \cdot \frac{\exp\left\{ -\frac{1}{2}\left(\frac{y-m_2}{\sigma_2}\right)^2 \right\}}{\sigma_2\sqrt{2\pi}} \\
&= f_X(x) f_Y(y)
\end{aligned}
$$

Where f_X, f_y are the pdf:s of two univariate normally distributed Gaussians $X \in N(m_1, \sigma_1^2)$, $Y \sim N(m_2, \sigma_2^2)$.

As we have $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, X and Y are independent.

In conclusion, we have shown that multivariate normally distributed Gaussians with correlations $\rho = 0$ are independent.

2

We have, under the NB assumption:

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k|Y) \\
&= P(X_1 = x_1|X_2 = x_2, \dots, X_k = x_k|Y) \cdot P(X_2 = x_2|X_1 = x_1, X_3 = x_3, \dots, X_k = x_k|Y) \\
&\cdots P(X_k = x_k|X_1 = x_2, X_2 = x_2, \dots, X_{k-1} = x_{k-1}|Y) \\
&= P(X_1 = x_1|Y) \cdot P(X_2 = x_2|Y) \cdots P(X_k = x_k|Y) \\
&= \prod_{i=1}^{k} P(X_i = x_i|Y) \\
&= \prod_{i=1}^{k} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2}\big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 \Big\}
\end{aligned}
$$

From the multivariate Gaussian distribution, we also have

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n|Y) \\
&= f_X(\bar{x}) = \frac{1}{(2\pi)^{k/2} \det \Sigma^{1/2}} \exp\Big\{ -\frac{1}{2}(\bar{x} - \mu_Y)^T \Sigma^{-1} (\bar{x} - \mu_Y)^T \Big\}
\end{aligned}
$$

Taking logs, we have that for equality to hold we need Σ to satisfy:

$$
\begin{aligned}
&\log \prod_{i=1}^{k} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2}\big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 \Big\} \\
&= \log \frac{1}{(2\pi)^{k/2} \det \Sigma^{1/2}} \exp\Big\{ -\frac{1}{2}(\bar{x} - \mu_Y)^T \Sigma^{-1} (\bar{x} - \mu_Y)^T \Big\} && \Leftrightarrow \\
&-\frac{1}{2}\sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 - \sum_{i=1}^{k} \log \sigma_i\sqrt{2\pi} \\
&= -\frac{1}{2}(\bar{x} - \mu_Y)^T \Sigma^{-1} (\bar{x} - \mu_Y)^T - \log\,[(2\pi)^{k/2} \det \Sigma^{1/2}] && \Leftrightarrow \\
&-\frac{1}{2}\sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 - \frac{1}{2}\sum_{i=1}^{k} \log \sigma_i^2 - \frac{1}{2}\sum_{i=1}^{k} \log 2\pi \\
&= -\frac{1}{2}(\bar{x} - \mu_Y)^T \Sigma^{-1} (\bar{x} - \mu_Y)^T - \frac{k}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma && \Leftrightarrow \\
&-\frac{1}{2}\Big[\sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 - \sum_{i=1,j=1}^{k} (x_i - \mu_{Y,i})^T (\Sigma^{-1})_{i,j} (x_j - \mu_{Y,j})^T\Big] \\
&= \frac{1}{2}\log\prod_{i=1}^{k} \sigma_i^2 - \frac{1}{2}\log\det\Sigma && \Leftrightarrow
\end{aligned}
$$

Comparing the non-constant expressions, which need to be equal for the equality to hold:

$$\begin{aligned}
\sum_{i=1,j=1}^{k} \big(x_i - \mu_{Y,i}\big)^T (\Sigma^{-1})_{i,j} \big(x_j - \mu_{Y,j}\big)^T &= \sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 \Leftrightarrow \\
\sum_{t=1}^{k} \big(x_t - \mu_{Y,t}\big)^T (\Sigma^{-1})_{t,t} \big(x_t - \mu_{Y,t}\big)^T + \sum_{i\neq j}^{k} \big(x_i - \mu_{Y,i}\big)^T (\Sigma^{-1})_{i,j} \big(x_j - \mu_{Y,j}\big)^T &= \sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 \Leftrightarrow \\
\underbrace{\sum_{t=1}^{k} \big(x_t - \mu_{Y,t}\big)^2 (\Sigma^{-1})_{t,t}}_{A} + \underbrace{\sum_{i\neq j}^{k} \big(x_i - \mu_{Y,i}\big) \big(x_j - \mu_{Y,j}\big) (\Sigma^{-1})_{i,j}}_{B} &= \sum_{i=1}^{k} \big(\frac{x_i - \mu_{Y,i}}{\sigma_i}\big)^2 \Leftrightarrow
\end{aligned}$$

As the second term B contains "cross-terms" $x_i x_j, i \neq j$, but the right hand side contains only quadratic terms, such as these in A, we have that B must be 0. As such $(\Sigma^{-1})_{i,j} = 0$ for $i \neq j$ $\implies \Sigma^{-1}$ diagonal $\implies \Sigma$ diagonal. For the equality to hold $(\Sigma^{-1})_{i,i} = \frac{1}{\sigma_i^2} \implies \Sigma_{i,i} = \sigma_i^2$. For this Σ, we can see that the contant terms are equal as well, as the determinant of a diagonal matrix is the product of its diagonal elements $\therefore \det \Sigma = \prod_{i=1}^{k} \sigma_i^2$.

3

3.1 (a)

We have

$$
\begin{aligned}
P(C_0|x) &= \frac{P(x|C_0)P(C_0)}{P(x|C_0)P(C_0) + P(x|C_1)P(C_1)} \\
&= \frac{1}{1 + \frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)}} \\
&= \sigma(\frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)})
\end{aligned}
$$

As such

$$
\begin{aligned}
\exp(-a) &= \frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)} \Leftrightarrow \\
a &= -\ln\frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)} \\
&= \ln\frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)}
\end{aligned}
$$

3.2 (b)

We get (using $a^T = a$ for scalar a, Σ symmetric $\implies x^T\Sigma^{-1}y = y^T(\Sigma^{-1})^Tx = y^T(\Sigma^T)^{-1}x = y^T\Sigma^{-1}x$):

$$
\begin{aligned}
a &= \ln \frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)} \\
&= \ln \frac{\frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\Big)P(C_0)}{\frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\Big)P(C_1)} \\
&= -\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) + \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \ln\frac{P(C_0)}{P(C_1)} \\
&= -\frac{1}{2}x^T\Sigma^{-1}x + \frac{1}{2}\mu_0^T\Sigma^{-1}x + \frac{1}{2}x^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 \\
&+ \frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}\mu_1^T\Sigma^{-1}x - \frac{1}{2}x^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\frac{P(C_0)}{P(C_1)} \\
&= \mu_0^T\Sigma^{-1}x - \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}x + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\frac{P(C_0)}{P(C_1)} \\
&= (\mu_0-\mu_1)^T\Sigma^{-1}x + \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_0^T\Sigma^{-1}\mu_0) + \ln\frac{P(C_0)}{P(C_1)}
\end{aligned}
$$

As such, we have a linear decision boundary, with $w = \Sigma^{-1}(\mu_0-\mu_1)$,
$b = \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_0^T\Sigma^{-1}\mu_0) + \ln\frac{P(C_0)}{P(C_1)}$

3.3 (c)

We get:

$$
\begin{aligned}
a &= \ln \frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)} \\
&= \ln \frac{\frac{1}{(2\pi)^{k/2}|\Sigma_0|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0)\Big)P(C_0)}{\frac{1}{(2\pi)^{k/2}|\Sigma_1|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\Big)P(C_1)} \\
&= -\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) + \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \underbrace{\ln\frac{P(C_0)|\Sigma_1|^{1/2}}{P(C_1)|\Sigma_0|^{1/2}}}_{=K} \\
&= -\frac{1}{2}x^T\Sigma_0^{-1}x + \frac{1}{2}\mu_0^T\Sigma_0^{-1}x + \frac{1}{2}x^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 + \frac{1}{2}x^T\Sigma_1^{-1}x - \frac{1}{2}\mu_1^T\Sigma_1^{-1}x - \frac{1}{2}x^T\Sigma_1^{-1}\mu_1 + \frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 + K \\
&= x^T\frac{1}{2}(\Sigma_1^{-1} - \Sigma_0^{-1})x + \mu_0^T\Sigma_0^{-1}x - \mu_1^T\Sigma_1^{-1}x - \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 + \frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 + K \\
&= x^T\frac{1}{2}(\Sigma_1^{-1} - \Sigma_0^{-1})x + (\Sigma_0^{-1}\mu_0 - \Sigma_1^{-1}\mu_1)^T x - \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 + \frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 + \ln\frac{P(C_0)|\Sigma_1|^{1/2}}{P(C_1)|\Sigma_0|^{1/2}}
\end{aligned}
$$

As such, we have a quadratic decision boundary, with $w = \Sigma_0^{-1}\mu_0 - \Sigma_1^{-1}\mu_1$,
$b = \frac{1}{2}(\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_0^T\Sigma_0^{-1}\mu_0) + \ln\frac{P(C_0)|\Sigma_1|^{1/2}}{P(C_1)|\Sigma_0|^{1/2}}$

4

4.1 (a)

Assuming independence between observations, we get:

$$
\begin{aligned}
&P(x^{(i)}, \dots, x^{(m)}, y^{(i)}, \dots, y^{(m)}) \\
&= \prod_{i=1}^{m} P(x^{(i)}, y^{(i)}) \\
&= \prod_{i=1}^{m} P(y^{(i)}) P(x^{(i)} | y^{(i)}) \\
&= \prod_{i=1}^{m} \Big[\frac{\phi}{(2\pi)^{k/2} |\Sigma_1|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1) \Big) \Big]^{\mathbb{1}(y^{(i)}=1)} \\
&\cdot \Big[\frac{1-\phi}{(2\pi)^{k/2} |\Sigma_0|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0) \Big) \Big]^{\mathbb{1}(y^{(i)}=0)}
\end{aligned}
$$

As such, we have:

$$
\begin{aligned}
L(\phi, \mu_0, \mu_1, \Sigma) &= \ln \prod_{i=1}^{m} P(x^{(i)} | y^{(i)}) P(y^{(i)}) \\
&= \ln \prod_{i=1}^{m} \Big[\frac{\phi}{(2\pi)^{k/2} |\Sigma_1|^{1/2}} \exp\Big(-\frac{1}{2}(x^{(i)}-\mu_1)^T \Sigma_1^{-1} (x^{(i)}-\mu_1) \Big) \Big]^{\mathbb{1}(y^{(i)}=1)} \\
&\cdot \Big[\frac{1-\phi}{(2\pi)^{k/2} |\Sigma_0|^{1/2}} \exp\Big(-\frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma_0^{-1} (x^{(i)}-\mu_0) \Big) \Big]^{\mathbb{1}(y^{(i)}=0)} \\
&= \sum_{i=1}^{m} \ln \frac{1}{(2\pi)^{k/2} |\Sigma_1|^{1/2}} + \mathbb{1}(y^{(i)} = 1) \Big[\ln \phi + \Big(-\frac{1}{2}(x^{(i)}-\mu_1)^T \Sigma_1^{-1} (x^{(i)}-\mu_1) \Big) \Big] \\
&+ \mathbb{1}(y^{(i)} = 0) \Big[\ln(1-\phi) + \Big(-\frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma_0^{-1} (x^{(i)}-\mu_0) \Big) \Big]
\end{aligned}
$$

4.2 (b)

Solving for a stationary point, only looking at the terms involving ϕ, we get:

$$
\begin{aligned}
\frac{\partial L}{\partial \phi} &= 0 \Leftrightarrow \\
\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1)\frac{1}{\phi} - \mathbb{1}(y^{(i)} = 0)\frac{1}{1-\phi} &= 0 \Leftrightarrow \\
\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1) - \phi \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1) - \phi \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) &= 0 \Leftrightarrow \\
\phi &= \frac{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1)}{m} \Leftrightarrow
\end{aligned}
$$

For this stationary point to be a maximum and hence the "best", $\frac{\partial^2 L}{\partial \phi^2} \leq 0$ for the point. We have:

$$
\frac{\partial^2 L}{\partial \phi^2} = \underbrace{-\underbrace{\frac{1}{\phi^2}}_{\geq 0} \underbrace{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1)}_{\geq 0} - \underbrace{\frac{1}{(1-\phi)^2}}_{\geq 0} \underbrace{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)}_{\geq 0}}_{\leq 0}
$$

4.3 (c)

Solving for a stationary point, only looking at the terms involving μ_0, we get:

$$
\begin{aligned}
\frac{\partial L}{\partial \mu_0} &= 0 \Leftrightarrow \\
\frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[(x^{(i)} - \mu_0)^T \Sigma_0^{-1} (x^{(i)} - \mu_0)\right] &= 0 \Leftrightarrow \\
\frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[{x^{(i)}}^T \Sigma_0^{-1} x^{(i)} - 2\mu_0^T \Sigma_0^{-1} x^{(i)} + \mu_0^T \Sigma_0^{-1} \mu_0\right] &= 0 \Leftrightarrow \\
\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[-2\Sigma_0^{-1} x^{(i)} + 2\Sigma_0^{-1} \mu_0\right] &= 0 \Leftrightarrow \\
\Sigma_0^{-1} \mu_0 \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) &= \Sigma_0^{-1} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) x^{(i)} \Leftrightarrow \\
\mu_0 &= \frac{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \mu_0} &= 0 \Leftrightarrow \\
\frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[(x^{(i)} - \mu_0)^T \Sigma_0^{-1} (x^{(i)} - \mu_0)\right] &= 0 \Leftrightarrow \\
\frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[{x^{(i)}}^T \Sigma_0^{-1} x^{(i)} - 2\mu_0^T \Sigma_0^{-1} x^{(i)} + \mu_0^T \Sigma_0^{-1} \mu_0\right] &= 0 \Leftrightarrow \\
&= 0 \Leftrightarrow \\
\Sigma_0^{-1} \mu_0 \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) &= \Sigma_0^{-1} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) x^{(i)} \Leftrightarrow \\
\mu_0 &= \frac{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0) x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)}
\end{aligned}
$$

For this stationary point to be a maximum and hence the "best", the hessian $\frac{\partial^2 L}{\partial \phi^2}$ must be negative definite. We have:

$$
\begin{aligned}
\frac{\partial^2 L}{\partial \phi^2} &= \frac{\partial L}{\partial \phi} \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)\left[\Sigma_0^{-1} x^{(i)} - \Sigma_0^{-1} \mu_0\right] \\
&= -\underbrace{[\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)]}_{>0} \Sigma_0^{-1}
\end{aligned}
$$

If a matrix A is positive definite, we have that $\alpha A, \alpha > 0$ is positive definite as well as $x^T \alpha A x = \underbrace{\alpha}_{>0} \underbrace{x^T A x}_{>0} \; \forall x \neq 0$.

As such, we have Σ_0^{-1} positive definite $\implies [\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)]\Sigma_0^{-1}$ positive definite $\implies \frac{\partial^2 L}{\partial \phi^2} = -[\sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 0)]\Sigma_0^{-1}$ negative definite $\implies \mu_0$ is the "best" estimate.

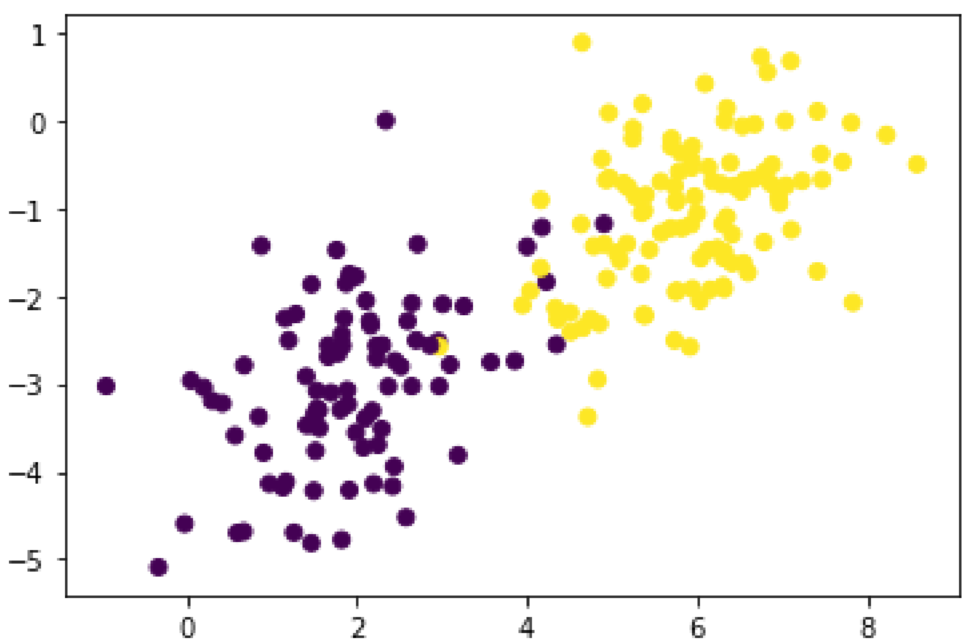
5

5.1 (a)

[23]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv("data.csv", header=None)

X = np.array(data[[0, 1]])
y = np.array(data[2])
```

[24]:
```
plt.scatter(X.T[0], X.T[1], c = y)
```

[24]:
```
<matplotlib.collections.PathCollection at 0x29557ab46c8>
```



No, the data is not linearly separable.

5.2 (b)

[25]:
```
def add_vectors(X):
    s = X[0]
    for i in range(1, X.shape[0]):
        s = np.add(s, X[i])
    return s

M1 = sum(y == 1)
X1 = X[y == 1]
mu_1 = 1/M1 * add_vectors(X1)

M0 = sum(y == 0)
X0 = X[y == 0]
mu_0 = 1/M0 * add_vectors(X0)

theta  = M1/len(y)

s1 = 1/M1 * sum([np.matmul(np.matrix(X1[i] - mu_1).T, np.matrix(X1[i] - mu_1))␣
 ↪for i in range(X1.shape[0])])
s0 = 1/M0 * sum([np.matmul(np.matrix(X0[i] - mu_0).T, np.matrix(X0[i] - mu_0))␣
 ↪for i in range(X0.shape[0])])
sigma = (s1*M1 + s0*M0) / (M1+M0)


print(f'theta = {theta}')
print(f'mu_0 = {mu_0}')
print(f'mu_1 = {mu_1}')
print(f'sigma = {sigma}')
```

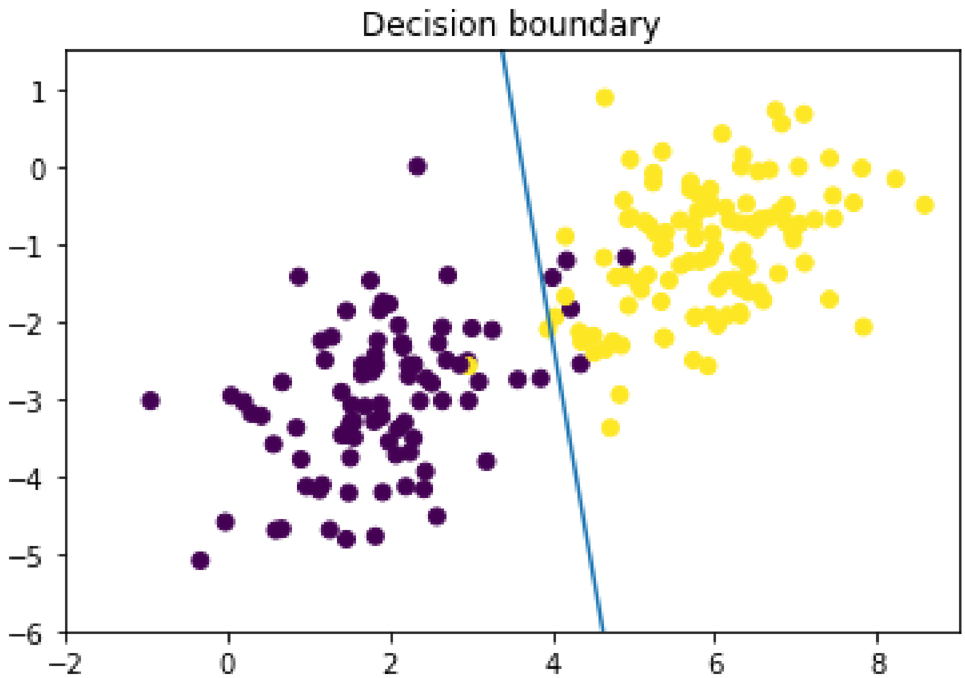
```
theta = 0.555
mu_0 = [ 1.9195151 -2.9972116]
mu_1 = [ 5.89818829 -1.07926031]
sigma = [[1.0180746  0.38866005]
 [0.38866005 0.80363858]]
```

5.3 (c)

[57]:
```
def quadratic_form(x, A):
    return x.dot(np.asarray(np.matmul(A, x)).reshape(-1))

# w = np.asarray(np.matmul(mu_0 - mu_1, np.linalg.inv(sigma))).reshape(-1)
# b = quadratic_form(mu_0 - mu_1, np.linalg.inv(sigma)) - np.log(M1/M0)

# print(f'w = {w}')
# print(f'b = {b}')

w = np.asarray((mu_0 - mu_1).dot(np.linalg.inv(sigma))).reshape(-1)
b = 1/2*(quadratic_form(mu_1, np.linalg.inv(sigma)) - quadratic_form(mu_0, np.
 ↪linalg.inv(sigma))) - np.log(M0/M1)


print(f'w = {w}')
print(f'b = {b}')
```

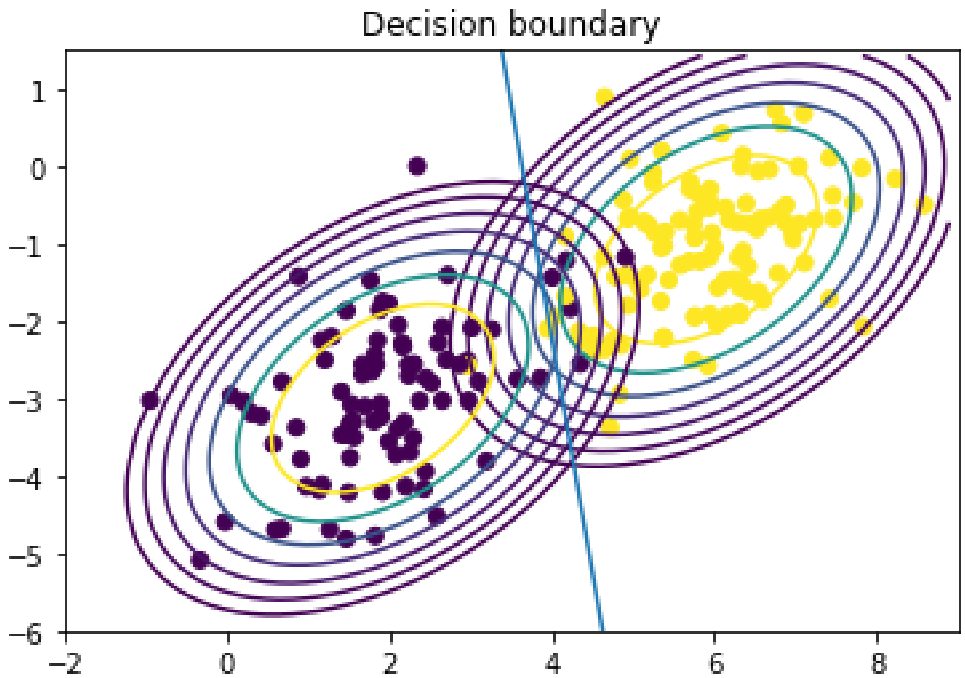
```
w = [-3.67554649 -0.60899668]
b = 13.346781014999657
```

[58]:
```
plt.scatter(X.T[0], X.T[1], c = y)

x1 = np.array([-2,9])
x2 = (-b - x1 * w[0]) / w[1]

plt.plot(x1, x2)
plt.xlim(-2,9)
plt.ylim(-6,1.5)
plt.title("Decision boundary")
plt.show()
```


Decision boundary
1
0
−1
−2
−3
−4
−5
−6
−2
0
2
4
6
8

5.4 (d)

[59]:
```
def gaussian(x, y, mu, sigma):
    C = 1/(2*np.pi)*np.sqrt(np.linalg.det(sigma))
    X = np.array([x, y])
    Z = quadratic_form(X - mu, np.linalg.inv(sigma))
    return C*np.exp(-1/2*Z)

def plot_contour(mu):
    nn = 100
    xx = np.arange(-2, 9, (14.5+2)/nn)
    yy = np.arange(-6, 1.5, (1.5+6)/nn)
    XX, YY = np.meshgrid(xx, yy)

    Z = np.empty(XX.shape)
    for i in range(XX.shape[0]):
        for j in range(XX.shape[1]):
            Z[i, j] = gaussian(XX[i, j], YY[i, j], mu, sigma)

    levels = 10**(np.arange(-3, -1, 2/7))
    plt.contour(XX, YY, Z, levels)
    plt.xlim(-2,14.5)
    plt.ylim(-6,1.5)


plot_contour(mu_1)
plot_contour(mu_0)

plt.scatter(X.T[0], X.T[1], c = y)

x1 = np.array([-2,9])
x2 = (-b - x1 * w[0]) / w[1]

plt.plot(x1, x2)
plt.xlim(-2,9)
plt.ylim(-6,1.5)
plt.title("Decision boundary")
plt.show()
```



The decision boundary does pass through the points where the distribution have equal probabilities, as these points are the points where the probabilities are equal, i.e. $P(Y = 1|X) = P(Y = 0|X)$, which is the definition of the decision boundary.