ComS 574 HW 3 Kanak Choudhury

Problem 1: Ch-4, Q7:

We have, $X \sim N(\mu_i, \sigma^2)$; $i = 1, 2$

where, μ_1 is the mean of last year's percent profit for companies that issued a dividend. and

μ_2 is the mean of last year's percent profit for companies that didn't issued a dividend.

we have, $\hat{\mu}_1 = \bar{x}_1 = 10$

$\hat{\mu}_2 = \bar{x}_2 = 0$

and $\hat{\sigma}^2 = 36$

Prior probability, $P(Yes) = 0.8$ & $P(No) = 0.2$

So, the posterior probability-

$$P(Yes \mid X=4) = \frac{0.8 \times \cancel{\frac{1}{\sqrt{2\pi 36}}} \exp\left[-\frac{1}{2\times 36}(4-10)^2\right]}{\cancel{\frac{1}{\sqrt{2\pi 36}}}\left\{0.8 \exp\left[-\frac{1}{2\times 36}(4-10)^2\right] + 0.2 \exp\left[-\frac{1}{2\times 36}(4-0)^2\right]\right\}}$$

$= 0.75185$

So, the probability that a company will issue a dividend this year given that its percentage profit was $X=4$ last year is 0.75185.

Problem 2:

We have,

$$\Sigma_{Y=+1} = \Sigma_{Y=+2} = \Sigma_{Y=+3} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Prior probability: $\pi_{Y=+1} = \pi_{Y=+2} = \pi_{Y=+3} = 1/3$

and $\mu_{Y=+1} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $\mu_{Y=+2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ & $\mu_{Y=+3} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

~~As~~ We know, the Bayes classifier assigns an observation $X = x$ to the class for which

$$\delta_k(x) = x'\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k'\Sigma^{-1}\mu_k + \log \pi_k$$

a) LDA boundary bet^n $Y = +1$ and $Y = +2 \Rightarrow$

$$\delta_1(x) = \delta_2(x)$$

$$\Rightarrow x'\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1'\Sigma^{-1}\mu_1 + \log \pi_1 = x'\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2'\Sigma^{-1}\mu_2 + \log \pi_2$$

Where, $x' = [x_1 \; x_2]$, $\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$\Rightarrow x'\begin{pmatrix} -1 \\ -1 \end{pmatrix} - \cancel{\frac{1}{2}[-1 \; -1]\begin{pmatrix} -1 \\ -1 \end{pmatrix}} + \cancel{\log \frac{1}{3}} = x'\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \cancel{\frac{1}{2}[1 \; 1]\begin{pmatrix} 1 \\ 1 \end{pmatrix}} + \cancel{\log \frac{1}{3}}$$

$$\Rightarrow -x_1 - x_2 = x_1 + x_2$$

$$\Rightarrow 2x_1 + 2x_2 = 0$$

$$\Rightarrow x_1 + x_2 = 0$$

b) LDA boundary betⁿ $Y=+1$ & $Y=+3$

$$\delta_1(x) = \delta_3(x)$$

$$\Rightarrow x'\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1'\Sigma^{-1}\mu_1 + \cancel{\log \pi_1} = x'\Sigma^{-1}\mu_3 - \frac{1}{2}\mu_3'\Sigma^{-1}\mu_3 + \cancel{\log \pi_3}$$

$$\Rightarrow -\cancel{x_1} - x_2 - \cancel{1} = -\cancel{x_1} + x_2 - \cancel{1}$$

$$\Rightarrow x_2 = 0$$

c) LDA boundary betⁿ $Y=+2$ & $Y=+3$

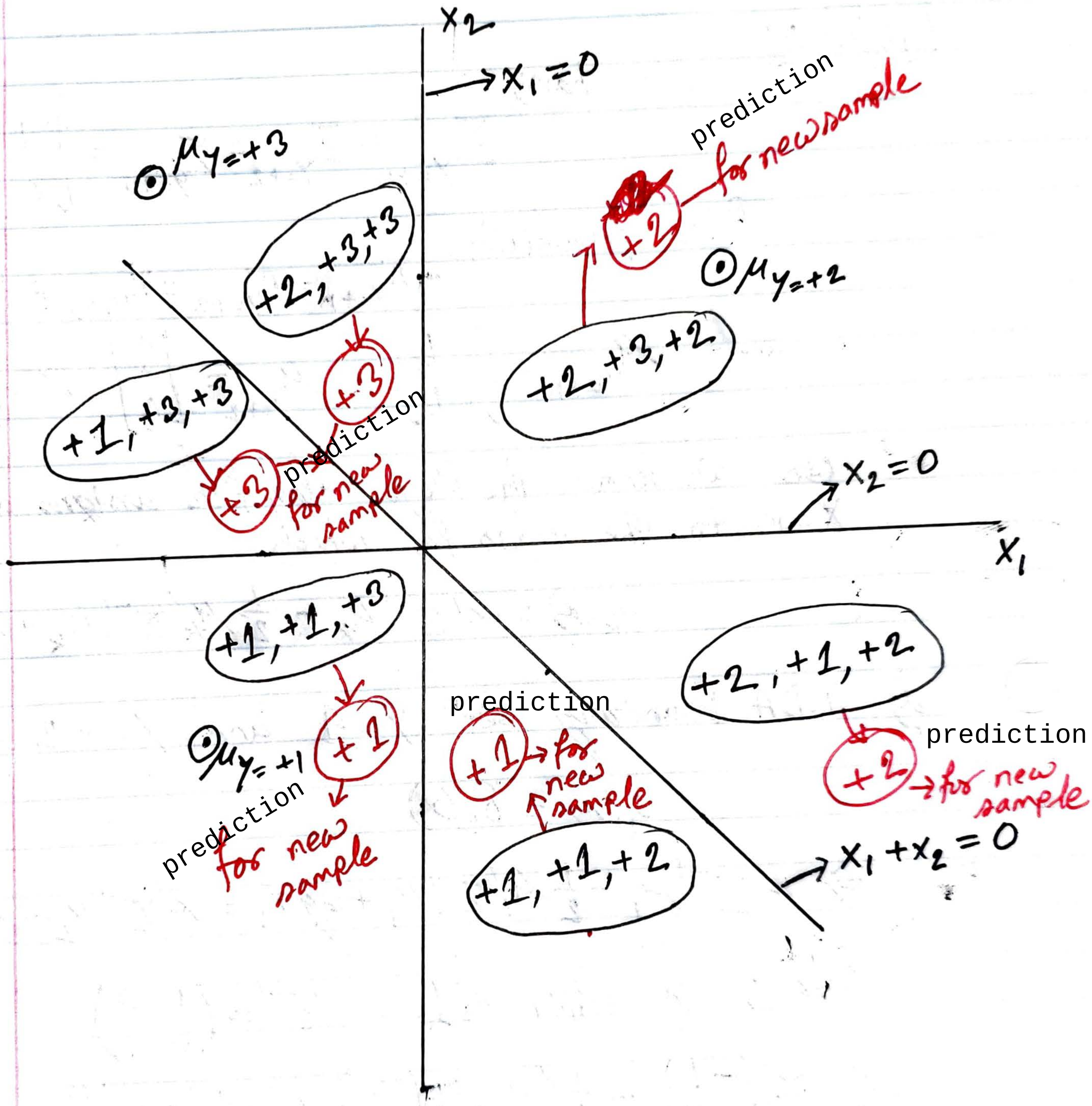
$$\delta_2(x) = \delta_3(x)$$

$$\Rightarrow x'\Sigma^{-1}\mu_2 - \cancel{\frac{1}{2}\mu_2'\Sigma^{-1}\mu_2} + \cancel{\log \pi_2} = x'\Sigma^{-1}\mu_3 - \cancel{\frac{1}{2}\mu_3'\Sigma^{-1}\mu_3} + \cancel{\log \pi_3}$$

$$\Rightarrow x_1 + \cancel{x_2} = -x_1 + \cancel{x_2}$$

$$\Rightarrow x_1 = 0$$

d)



Based on majority voting—

e) For this exercise, since the covariance matrix is same for all three groups, there will be no change even if we try to use QDA.

Because, in QDA decision function,

$$\delta_k(x) = -\frac{1}{2}x'\Sigma_k^{-1}x + x'\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k + \log\pi_k$$

$\Sigma_k = \Sigma \Rightarrow$ the extra term $-\frac{1}{2}x'\Sigma_k^{-1}x$ is same for all classes and it will cross out from the boundary equation.

f) For the naive Bayes classification, we consider that the X's are independent random variables. that is, $cov(\underset{\sim}{X}_i, \underset{\sim}{X}_j) = 0$.

For this problem X_1 and X_2 are independent. as a result, both LDA and naive Bayes will produce same results for this problem.

Thus, there will be no change if we use LDA and naive Bayes separately.