Marco Reina 7066486
Samuele Serri 7069839

a)

for class y=0
$E[x1] = (1 + 1 + 2 + 3 + 3)/5 = 2$
$E[x2] = (1 + 1 + 2 + 2 + 3)/5 = 1.8$
$cov(x1, x2) = E[x1x2] - E[x1]E[x2] = (3 + 6 + 6 + 1 + 2)/5 - 3.6 = 0$
for class y=1:
$E[x1] = (1 + 2 + 4 + 5 + 5)/5 = 3.4$
$E[x2] = (4 + 5 + 6 + 6 + 7)/5 = 5.6$
$cov(x1, x2) = E[x1x2] - E[x1]E[x2] = (30 + 24 + 20 + 10 + 7)/5 - 19.04 = 18.2 - 19.04 = -1.05$

b)

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

$x = (3.5, 2)^T$,
$\mu_0 = (2, 1.8)^T$,
$\mu_1 = (3.4, 5.6)^T$,
$\Sigma = \frac{1}{8}(C_0^T - \mu_0)(C_1^T - \mu_1)^T = \frac{1}{8} \begin{pmatrix} 8 & -4.2 \\ -4.2 & 17.2 \end{pmatrix}$.

Where $C_1 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \\ 2 & 3 \\ 1 & 3 \end{pmatrix}$ and $C_2 = \begin{pmatrix} 4 & 5 \\ 6 & 5 \\ 6 & 4 \\ 5 & 2 \\ 7 & 1 \end{pmatrix}$ And we compute:

$\delta_0 = 7.7027$
$\delta_1 = 5.4846$
With this results we classify the point x as a class 0 point.

c)

LDA assumes that the classes have different means and shared variance, while with QDA each class can have a different variance. Both assume a gaussian distribution.

d)

e)

LDA is a much less flexible classifier than QDA. therefore LDA usually makes better predictions when there are relatively few training observations and reducing variance is crucial. QDA can be used with a bigger sample size, when the variance of the classifier is not a huge concern.