Ex3:

a) Class 0 ($y=0$):

mean:

$$\mu_0 = \frac{1}{5}\sum_{i=1}^{5} x_i = \left(\frac{1+1+2+2+3}{5}, \frac{1+1+2+3+3}{5}\right) = (1{,}8, 2)$$

$\bar{x}_0 = 1{,}8 \quad \bar{y}_0 = 2$

$$\mathrm{Cov}(x_0, y_0) = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{x}_0)(y_i - \bar{y}_0) = \frac{1}{4}(0{,}8 - 0{,}2 + 0{,}2 - 0{,}8) = 0$$

Class 1 ($y=1$):

mean:

$$\mu_1 = \frac{1}{5}\sum_{i=1}^{5} x_i = \left(\frac{4+5+6+6+7}{5}, \frac{1+2+4+5+5}{5}\right) = (5{,}6, 3{,}4)$$

$\bar{x}_1 = 5{,}6 \quad \bar{y}_1 = 3{,}4$

$$\mathrm{Cov}(x_1, y_1) = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{x}_1)(y_i - \bar{y}_1) = \frac{1}{4}(-3{,}36 + 0{,}84 + 0{,}24 + 0{,}64 - 2{,}56)$$

$$= \frac{1}{4}\cdot(-4{,}2)$$

$$= -1{,}0\overline{5}$$

b) $\bar{x} = \frac{1}{10}(1+1+2+2+3+4+5+6+6+7) = 3{,}7$

$\bar{y} = \frac{1}{10}(1+1+1+2+2+3+3+4+5+5) = 2{,}7$

$$\sigma(x,x) = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2 = \frac{1}{9}(2\cdot 7{,}29 + 2\cdot 2{,}89 + 0{,}49 + 0{,}09 + 1{,}69 + 2\cdot 5{,}29 + 10{,}89)$$

$$= 4{,}9$$

$$\sigma(y,y) = \frac{1}{9}\sum_{i=1}^{10}(y_i - \bar{y})^2 = \frac{1}{9}(3\cdot 2{,}89 + 2\cdot 0{,}49 + 2\cdot 0{,}09 + 1{,}69 + 2\cdot 5{,}29)$$

$$= 2{,}4\overline{5} \approx 2{,}46$$

$$\sigma(x,y) = \sigma(y,x) = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{9}(4{,}59 + 2{,}89 - 5{,}61 + 0{,}49 - 0{,}91 - 0{,}81 - 0{,}51 + 2{,}59 + 0{,}69 + 5{,}29)$$

$$= 1{,}0\overline{1} \approx 1{,}01$$

$$\Sigma = \begin{bmatrix} 4{,}9 & 1{,}0\overline{1} \\ 1{,}0\overline{1} & 2{,}4\overline{5} \end{bmatrix} \qquad \Sigma^{-1} \approx \frac{1}{\underbrace{11{,}03}_{\det(\Sigma)}} \begin{bmatrix} 2{,}46 & -1{,}01 \\ -1{,}01 & 4{,}9 \end{bmatrix} = \begin{bmatrix} 0{,}22 & -0{,}09 \\ -0{,}09 & 0{,}44 \end{bmatrix}$$

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

$$\delta_1(x) = (3{,}5, 2)\,\Sigma^{-1}\begin{pmatrix}1{,}8\\2\end{pmatrix} - \frac{1}{2}(1{,}8, 2)\,\Sigma^{-1}\begin{pmatrix}1{,}8\\2\end{pmatrix} + \log(0{,}5)$$

$$= 2{,}192 - \frac{1}{2}\cdot 1{,}9248 + \log(0{,}5)$$

$$\approx 0{,}586\overline{5}$$

$$\delta_2(x) = (3{,}5, 2)\,\Sigma^{-1}\begin{pmatrix}5{,}6\\3{,}4\end{pmatrix} - \frac{1}{2}(5{,}6, 3{,}4)\,\Sigma^{-1}\begin{pmatrix}5{,}6\\3{,}4\end{pmatrix} + \log(0{,}5)$$

$$= 5{,}225 - \frac{1}{2}\cdot 8{,}5584 + \log(0{,}5)$$

$$\approx 0{,}2527$$

c)

LDA

- it is assumed, that all classes share one covariance matrix.
- each class is gaussian distributed
- class boundaries are linear

QDA:

- each class has its own covariance matrix.
- each class is gaussian distributed
- class boundaries are quadratic

d) Since we assume, that each class is gaussian distributed, the likelihood of a point belonging to a class is proportional to the distance to the mean of the class.
Thus even if the input has n dimensions, for deciding between two output classes, one dimension is enough.
The minimum dimensionality while maintaining class separability is $c-1$, with c being the number of classes.

e) When to choose LDA:

- Sample Size: When the sample size is small, since LDA has fewer parameters and is thus less prone to overfitting.
- number of features: High number of features, since LDA scales better with its simpler covariance.
- model complexity: Less complex models with linear decision boundarys.

When to choose QDA:

- Sample Size: Large sample size, since QDA has separate covariance matrices, which need more data to estimate the parameters
- number of features: low number of features, as a high number of features requires significantly more parameters.
- model complexity: high complexity, as it can adjust better to complex relations, with its quadrati decision boundarys.