1-1

1. Priors: π_k $\quad x = N(\mu, \sigma^2)$

$$P_x(k) = P(x \mid y=k)\,P(y=k) =$$

$$\pi_k \cdot \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}$$

Does not depend on k so not really necessary for calculation - can be removed

$$\log P_x(k) = \log(\pi_k) - \log(\sqrt{2\pi}\,\sigma_k) - \frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2$$

$$= \log(\pi_k) - \frac{1}{2\sigma^2}\left[x^2 - 2x\mu_k + \mu_k^2\right]$$

$$= \log(\pi_k) - \frac{x^2}{2\sigma^2} + \frac{x\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}$$

Again, this does not depend on k so it can be removed

$$\boxed{P_x(k) = x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)}$$

2. This is the LDA classifier.

1-2

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

2. No, the relationship is quadratic with respect to the feature vector x.

3. QDA classifier.

1-3

LDA is a simpler estimator, which means lower variance and higher bias. QDA is more complex, meaning it is more susceptible to high variance and overfitting.