Bernhard Nordemann
CS 4342

1. .

(1) $$P(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$P(x)\left(1 + e^{\beta_0 + \beta_1 x}\right) = e^{\beta_0 + \beta_1 x}$$

$$P(x) + P(x)\, e^{\beta_0 + \beta_1 x} = e^{\beta_0 + \beta_1 x}$$

$$\frac{P(x)}{e^{\beta_0 + \beta_1 x}} + \frac{\cancel{e^{\beta_0 + \beta_1 x}}}{P(x)} = 1$$

$$\frac{P(x)}{e^{\beta_0 + \beta_1 x}} = 1 - P(x)$$

$$\frac{1}{e^{\beta_0 + \beta_1 x}} = \frac{1 - P(x)}{P(x)}$$

$$e^{\beta_0 + \beta_1 x} = \frac{P(x)}{1 - P(x)} \rightarrow \boxed{\frac{P(x)}{1 - P(x)} = e^{\beta_0 + \beta_1 x}}$$

2. .

(2) we can ignore the denominator of $P_K(x)$ as it is same for all, so

$$f'(x) = \pi_K \frac{1}{\sqrt{2\pi\sigma}} e^{(-1/2\sigma^2 (x-\mu_K)^2)}$$

take natural log:

$$f''(x) = \ln \pi_K + \ln\left(\frac{1}{\sqrt{2\pi\sigma}}\right) - 1/2\sigma^2 (x-\mu_x)^2$$

2nd term is constant, so remove:

$$f'''(x) = \ln \pi_K - \frac{1}{2\sigma^2}(x + \mu_K)^2$$

$$f'''(x) = \ln \pi_K - 1/2\sigma^2 (x^2 - 2x\mu_K + \mu_K^2)$$

$$f'''(x) = \ln \pi_K - \frac{x^2}{2\sigma^2} + \frac{2x\mu_K}{2\sigma^2} - \frac{\mu_K^2}{2\sigma^2}$$

Remove the 2nd term, as it is constant for all.

$$f''''(x) = \ln \pi_K + \frac{x\mu_K}{\sigma^2} - \frac{\mu_K^2}{2\sigma^2}$$

we can rearrange to look like $\delta_K(x)$

$$f''''(x) = \delta_K(x) = x\frac{\mu_K}{\sigma^2} - \frac{\mu_K^2}{2\sigma^2} + \ln \pi_K$$

So, if you ignore all constants (values that will not affect whether a class will be a maximum), you get the elements which leverage max probability.

3. .

(3) continuing from $f''(x)$ in the last question (a.2)

$$f''(x) = \ln \pi_K + \ln\left(\frac{1}{\sqrt{2\pi}\sigma_K}\right) + -\frac{1}{2\sigma_K^2}(x-\mu_K)^2$$

$$f''(x) = \ln\left(\frac{\pi_K}{\sqrt{2\pi}\sigma_K}\right) - \left(x^2 - 2x\mu_K + \mu_K^2\right)/(2\sigma_K^2)$$

The first term is a constant for each class, and the second term is quadratic.

4. .
   a. We expect QDA to perform better on the training set. QDA has more flexibility so it performs better on the train data. We expect LDA to perform better on the test set, since if the Bayes decision boundary is linear, a linear model will be closer to what is the true boundary.
   b. We expect QDA to perform better on both.
   c. We expect the prediction of QDA relative to LDA to improve, as a larger training data set will be beneficial to QDA, as it is a more flexible model.
   d. This is False. In the case that there are few data points, QDA will be heavily affected by the variance of the data, while LDA will fit better. Also, it is easier for LDA to model a linear decision boundary than QDA, so if the Bayes decision boundary is linear, then the test data for LDA will likely be superior.

5. .

(5)

$$Y = \frac{e^{-6+0.05X_1+X_2}}{1+e^{-6+0.05X_1+X_2}}$$

a) $x_1 = 40h$ $\quad X_2 = 3.5$

$$e^{-6+0.05(40)+3.5}/(1+e^{-6+0.05(40)+3.5}) = 0.37754$$

b) set exponent of e to $\emptyset$

$$-6+0.05(X_1)+3.5 = 0$$

$X_1 = 50$ hours.

6. .

⑥ we can use $P_K(x)$

$$P_1(4) = \frac{0.8e^{-(1/72)(4-10)^2}}{0.8e^{-(1/72)(4-10)^2} + 0.2e^{-(1/72)(4-0)^2}} = 0.752$$

$P(x=4) = 0.752$

7. If we have a K value of K=1 on train data, we will have a near 0% error rate when testing against the train data. This means that if we are taking an average of the test and training data sets, we can calculate the test data error to be nearly 36%, which is greater than the test data error for logistic regression. Because it has a lower error, the logistic regression would be preferred for classification of new observations.