Alexa Summers. Santhoshini Sree Bolisetty, Gireesh Kumar Muppalla
Homework #2
CS 5565 Dr. Song

1. $$P(X) = \frac{e^{B_0 + B_1 X}}{1 + e^{B0 - B_1 X}} \longrightarrow e^{B_0 - B_1 X} = ?$$

$$P = \frac{T}{1+T} \Rightarrow T = P + PT$$

$$T - PT = P \qquad (1 - P)T = P$$

$$T = \frac{P}{1-T} = 1 + e^{B0 - B_1 X} =$$

$$\boxed{\frac{P(X)}{1 - P(X)} = e^{B_0 + B_1 X}}$$

5a. If linear, QDA should perform better because it is more flexible. On the test set, LDA should perform better because the QDA might overfit.

5b. LDA on the training set. LDA on the test set.

5c. Improvement would be expected because a more flexible method will have a better fit as the size increases.

5d. False. The variance from a more flexible method will cause an overfit.

6a.

$$P(x) = \frac{e^{B_0 + B_1X_1 + B_2X_2}}{1+e^{B_0+B_1X_1+B_2X_2}}$$

$$P(x) = \frac{e^{-6+.05X_1+X_2}}{1+e^{-6+.05X_1+X_2}}$$

$$P(x) = \frac{e^{-6+.05(.0540)+3.5}}{1+e^{-6+.05(.0540)+3.5}} = \frac{e^{-.05}}{1+e^{-.05}} = \boxed{37.75\%}$$

6b. $$P(x) = \frac{e^{-6+.05X_1+X_2}}{1+e^{-6+.05X_1+X_2}}$$

$$.50 = \frac{e^{-6+.05X_1+3.5}}{1+e^{-6+.05X_1+3.5}} \Rightarrow .50(1+e^{-2.5+.05X_1}) = e^{-2.5+.05X_1}$$

$$.50+.50e^{-2.5+.05X_1} = e^{-2.5+.05X_1}$$

$$.05 = .05e^{-2.5+.05X_1}$$

$$1 = -2.5+.05X_1$$

$$+2.5 \quad +2.5 \quad \frac{}{.05}$$

$$\frac{+2.5}{.05}$$

$$X_1 = 2.5/.05 = \boxed{50 \text{ hours}}$$

7. $$P_k(X) = \pi_k \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(X-M_k)^2}$$

$$\overline{\sum \pi_l \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\sigma^2(X-M_k)^2}}$$

$$P_{yes}(X) = \frac{\pi_{yes} e^{-1/2\sigma^2(X-M_{yes})^2}}{\sum \pi_l e^{-1/2\sigma^2(X-M_{yes})^2}}$$

$$= \frac{\pi_{yes} e^{-1/2\pi\sigma^2(X-M_{yes})^2}}{\pi_{yes} e^{-1/2\pi\sigma^2(X-M_{yes}^2)} + \pi_{no} e^{-1/2\pi\sigma^2(X-M_{no})^2}}$$

$$= .80e^{-1/2\cdot36(X-10)^2} / .80e^{-1/2\cdot36(X-10)^2} + .20e^{-1/2\cdot36X^2}$$

$$P_{yes}(4) = \frac{.80e(-1/2\cdot36(X-10)^2) + .20e^{-1/2\cdot36X^2}}{.80e(-1/2\cdot36(4-10)^2) + .20e^{-1/2\cdot36(16)}} = \boxed{75.2\%}$$

8. KNN has a test error rate of 36%, so logistic regression would be a better option because it's test error rate is 30%.

9a.

$$\frac{P(X)}{1-P(X)} = .37 \qquad P(X) = .37(1-P(X))$$

$$1.37p(X) = .37$$

$$P(X) = .37/1.37 = \boxed{27\%}$$

9b.

$$\text{defaultOdds} = \frac{P(X)}{1-P(X)} = .16/.84 = \boxed{0.19}$$

7 (LDA Problem)—see separate Excel sheet.