10/10/19.

VIGNESH PRABHAKAR
USC ID: 4024890667

Homework: 3

3. ~~7~~.4

(a) The training RSS of the polynomial regression would be lesser than the training RSS of the linear regression model because a higher degree polynomial would overfit the training data thereby reducing the RSS when compared to a linear model.

(b) Since the higher degree polynomial regression overfits the training data; it would ~~thus~~ thereby result in poor generalizations on the test data and thus resulting in a higher RSS_{test}. Whereas the linear model would generate a good fit that will also consider the generalizations and would therefore give lesser RSS_{test} on the test data.

(c) Polynomial regression will give us a lower training RSS in this case due to higher flexibility in modelling the ground truth which is non linear. therefore due to a higher flexibility; the polynomial regression model would always lower the training RSS more than the linear regression model irrespective of the degree of the underlying relationship between the predictor and the response variable

(d). We cannot say for sure and more details on the exact underlying relationship between the predictor and the response would be needed. Since it is mentioned that we dont know how far is the true relationship from linear; we cannot say for sure as to which model would lower the RSS more.
If the cubic regression is closer to the exact underlying relationship then a lower RSS would be observed with the cubic relationship than the linear relationship on the test data.
Otherwise if the linear regression is closer to the exact underlying relationship then a lower RSS would be observed with the linear relationship than the cubic relationship on the test data.

4.7.3

Assumption:

The Baye's formula assuming gaussianity among the classes and different class-variances is given by:

$$p_K(x) = \frac{\pi_K \frac{1}{\sqrt{2\pi}\sigma_K} \exp\left(-\frac{1}{2\sigma_K^2}(x-\mu_K)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma_K} \exp\left(-\frac{1}{2\sigma_l^2}(x-\mu_l)^2\right)}$$

Classify to the class with $\max_K p_K(x)$.

This is eqt. to finding K for which numerator is maximum.

$$K^* = \operatorname{argmax} \pi_K \frac{1}{\sqrt{2\pi}\sigma_K} \exp\left(-\frac{1}{2\sigma_K^2}(x-\mu_K)^2\right)$$

$$\log K^* = \operatorname{argmax} \log\left(\pi_K \frac{1}{\sqrt{2\pi}\sigma_K} e^{-\frac{1}{2\sigma_K^2}(x-\mu_K)^2}\right)$$

$$= \log \pi_K - \log\sqrt{2\pi} - \log \sigma_K - \frac{1}{2\sigma_K^2}(x-\mu_K)^2$$

$\log\sqrt{2\pi}$ is a constant & doesn't matter in maximization

$$= \arg\max \; \underbrace{\log \pi_K - \frac{1}{2}\left(\frac{x^2}{\sigma_K^2} - \frac{2\mu_K x}{\sigma_K} + \frac{\mu_K^2}{\sigma_K^2}\right) - \log \sigma_K}_{\delta(x)}$$

$\therefore$ discriminant score $= \delta(x)$

$$= \boxed{\log \pi_K - \log \sigma_K - \frac{1}{2}\left(\frac{x^2}{\sigma_K^2} - \frac{2\mu_K x}{\sigma_K} + \frac{\mu_K^2}{\sigma_K^2}\right)}$$

which is a quadratic function of x.

Thus the discriminant score ($\delta(x)$) is an quadratic function of x and the bayesian classifier is not linear.

Hence proved.

4.7.7

$$p_k(x) = \frac{\pi_k \cancel{\frac{1}{\sqrt{2\pi}\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum \pi_l \cancel{\frac{1}{\sqrt{2\pi}\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

$$= \frac{\pi_y \exp\left(-\frac{1}{2\sigma^2}(x-\mu_y)^2\right)}{\sum \pi_l \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

$$= \frac{0.80 \times \exp\left(-\frac{1}{2\times 36}(x-10)^2\right)}{0.80 \times \exp\left(-\frac{1}{2\times 36}(x-10)^2\right) + 0.20 \exp\left(-\frac{1}{2\times 36}(x-0)^2\right)}$$

$$p_k(4) = \frac{0.80 \times e^{-\frac{1}{\cancel{72}\,2} \times \cancel{36}}}{0.80 \times e^{-\frac{1}{2}} + 0.20 \times e^{-\frac{2}{9}}} = \frac{0.485}{0.485 + 0.160} = \boxed{75.2}$$