Question 2 : (ISLR 3.7.4)

a) It is difficult to say, we need more information for this. I would expect the polynomial regression to have a lower training RSS than the linear regression because it could make a tighter fit against data that matched with a wider irreducible error (Var ε). Train RSS for linear model are expected to be lower than the cubic regressor.

b) We need more information on the test set. I would expect the polynomial regression to have a higher test RSS as the overfit from training would have more error than the linear regression.

c) Polynomial regression has lower train RSS than the linear fit because of higher flexibility: no matter what the underlying true relationship is the more flexible model will closer follow points and reduce train RSS.

d) There is more information required to tell which will be lower. We do not know how far away from linear the true nature of the model is. If its closer to linear than cubic, the LR test RSS could be lower. Or, if it is closer to cubic than linear, the cubic regression test RSS could be lower. Unless the model is clearly specified it is difficult to say which test RSS is lower.

Question 3: (ISLR 4.7.3)

$$X \in N(\mu_k, \sigma_k)$$

Baye's Theorem:

$$p(x_k) = \frac{\pi_k f(x_k)}{\sum_{k'=1}^{K} \pi_{k'} f(x_{k'})} \quad \text{— (1)}$$

For normal distribution,

$$f(x_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-\frac{(x-\mu_k)^2}{2\sigma^2}} \quad \text{— (2)}$$

$$p(x_k) = \frac{\frac{\pi_k}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu_k)^2}{2\sigma^2}}}{\sum_{k'=1}^{K} \frac{\pi_{k'}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu_{k'})^2}{2\sigma^2}}} \quad \text{— From (1) \& (2)}$$

To maximize this we need large value of k,
Taking ln on both sides,

$$\ln(P(x_k)) = \ln\left[\frac{\pi_k}{\sqrt{2\pi\sigma^2}} \times e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}\right] - \underbrace{\ln\left[\sum_{k'=1}^{K} \frac{\pi_{k'}}{\sqrt{2\pi\sigma_{k'}^2}} \times e^{\frac{-(x-\mu_{k'})^2}{2\sigma^2}}\right]}$$

This term will be a constant for all k's. Thus it does not influence k; and thus we can ignore it.

Thus we get,

$$\ln(P(x_k)) = \ln \pi_k + \underbrace{\ln \frac{1}{\sqrt{2\pi\sigma^2}}} - \frac{(x-\mu_k)^2}{2\sigma^2}$$

Again this term $\frac{1}{\sqrt{2\pi}}$ doesn't contribute to $\underline{k}$.

So, we ignore it.

$$\delta(x_k) = \ln \pi_k - \ln \sigma_k - \frac{1}{2}\left(\frac{x^2}{\sigma_k^2} - \frac{2\mu_k x}{\sigma_k} + \frac{\mu_k^2}{\sigma_k^2}\right)$$

$$\delta(x_k) = ax^2 + bx + c$$

where

$$a = \frac{-1}{2\sigma_k^2}$$

$$b = \frac{-2\mu_k}{\sigma_k}$$

$$c = \ln \pi_k - \ln \sigma_k - \frac{1}{2}\frac{\mu_k^2}{\sigma_k^2}$$

It is of the form Quadratic equation.
Thus, Naive Bayes' classifier here is quadratic for $X \sim N(\mu_k, \sigma_k)$.

<u>Question 4:</u> (ISLR 4.7.7)

Here $k=2$ (Yes or No)

Let $k=1$ signifies Yes
$k=2$ signfies No

$\bar{X}$ for $k=1$ $=10$ , $\mu_1 = 10$
$\mu_2 = 0$

$\hat{\sigma}_2 = 36$ , $x = 4$

$\pi_1 = \frac{80}{100}$ , $\pi_2 = \frac{20}{100}$

$$P(x_k) = \frac{\pi_k \cdot f(x_k)}{\sum_{l=1}^{k} \pi_l \cdot f(x_l)}$$

So, $$f_k(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}$$

Thus plugging in the values, we get

$$P(k=1|x=4) = \frac{0.8 \times \cancel{\frac{1}{\sqrt{2\pi \cdot 36}}} \, e^{-\frac{(4-10)^2}{2\times 36}}}{0.8 \times \cancel{\frac{1}{\sqrt{2\pi \times 36}}} \times e^{-\frac{(4-10)^2}{2\times 36}} + 0.2 \times \cancel{\frac{1}{\sqrt{2\pi \times 36}}} \times e^{-\frac{(4-0)^2}{2\times 36}}}$$

$$= \frac{0.8 \times 0.6065}{0.8 \times 0.6065 + 0.2 \times 0.801} = 0.7519$$

$\Rightarrow$ $P(k=1|x=4) = 0.7519$ is the probability.