ASSIGNMENT 3 1666880

(i) (a) $\begin{cases} x_{11} = x_{12} = -x_{21} = -x_{22} \\ y_1 = -y_2 \end{cases}$

So $y = \begin{bmatrix} y_1 \\ -y_1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{11} \\ 1 & -x_{11} & -x_{11} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

$$\hat{\beta}_R = \arg\min_{\beta \in \mathbb{R}^3} \left[\sum_{i=1}^{2} (y_i - x_i\beta)^2 + \lambda \sum_{j=1}^{2} \beta_j^2 \right]$$

$$= \arg\min_{\beta \in \mathbb{R}^3} \left[(y_1 - \beta_0 - \beta_1 x_{11} - \beta_2 x_{11})^2 + (-y_1 - \beta_0 + \beta_1 x_{11} + \beta_2 x_{11})^2 + \lambda(\beta_1^2 + \beta_2^2) \right]$$

Use the fact that $(-a-b)^2 + (a-b)^2$ with $a = y_1 - \beta_1 x_{11} - \beta_2 x_{11}$ and $b = \beta_0$ to obtain:

$$\hat{\beta}_R = \arg\min_{\beta \in \mathbb{R}^3} \left[2(y_1 - \beta_1 x_{11} - \beta_2 x_{11})^2 + 2\beta_0^2 + \lambda \beta_1^2 + \lambda \beta_2^2 \right]$$

(b) Let $\Omega = 2(y_1 - \beta_1 x_{11} - \beta_2 x_{11})^2 + 2\beta_0^2 + \lambda\beta_1^2 + \lambda\beta_2^2$

Ω minimised at $\frac{\partial \Omega}{\partial \beta_j} = 0 \ \forall j = 0, 1, 2$.

$\Rightarrow \frac{\partial \Omega}{\partial \beta_1} = \frac{\partial \Omega}{\partial \beta_2}$, so,

$$-4x_{11}(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{11}) + 2\lambda\hat{\beta}_1 = -4x_{11}(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{11}) + 2\lambda\hat{\beta}_2$$

$$\hat{\beta}_1(4x_{11}^2 + 2\lambda) + \hat{\beta}_2(4x_{11}^2) = \hat{\beta}_1(4x_{11}^2) + \hat{\beta}_2(4x_{11}^2 + 2\lambda)$$

$$2\lambda\hat{\beta}_1 = 2\lambda\hat{\beta}_2$$

$$\hat{\beta}_1 = \hat{\beta}_2$$ as required.

(c) $$\hat{\beta}_L = \arg\min_{\beta \in \mathbb{R}^3} \left[\sum_{i=1}^{2} (y_i - x_i\beta)^2 + \sum_{j=1}^{2} |\beta_j| \right]$$

$$= \arg\min_{\beta \in \mathbb{R}^3} \left[2(y_1 - \beta_1 x_{11} - \beta_2 x_{11})^2 + 2\beta_0^2 + \lambda|\beta_1| + \lambda|\beta_2| \right]$$

(d) Let $\Omega = 2(y_1 - \beta_1 x_{11} - \beta_2 x_{11})^2 + 2\beta_0^2 + \lambda|\beta_1| + \lambda|\beta_2|$

Ω minimised at $\frac{\partial \Omega}{\partial \beta_j} = 0 \ \forall j = 0, 1, 2$.

$\Rightarrow \frac{\partial \Omega}{\partial \beta_1} = \frac{\partial \Omega}{\partial \beta_2}$, so $-4x_{11}(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{11}) + \lambda S(\beta_1) = -4x_{11}(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{11}) + \lambda S(\beta_2)$

where $S(z) = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \end{cases}$.

Now we have that,

$$\hat{\beta}_1(4x_{11}^2) + \hat{\beta}_2(4x_{11}^2) + \lambda S(\beta_1) = \hat{\beta}_1(4x_{11}^2) + \hat{\beta}_2(4x_{11}^2) + \lambda S(\beta_2)$$

which simplifies to $S(B_1) = S(B_2)$ for which there are infinitely many solutions ~~and hence $\hat{B}_1$ and $\hat{B}_2$ are not unique~~ and doesn't depend on the values of $\hat{B}_1$ and $\hat{B}_2$ and hence $\hat{B}_1$ and $\hat{B}_2$ are not unique.

(2) For least squares linear regression we have that

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_i} \right)$$

where $\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$, $\hat{B}_0, \hat{B}_1$ are the least square estimators of B_0 and B_1. And $h_i = h_{ii} = [H]_{ii}$ where

$H = X(X^T X)^{-1} X^T$ which in simple linear regression is simply

$$h_i = \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \right]$$

Hence
$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{y_i - (\hat{B}_0 + \hat{B}_1 x_i)}{1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \right)} \right]$$

3.a. The estimated standard errors for the coefficients of income and balance from the glm model are 4.985*10-6 and 2.274*10-4 respectively.

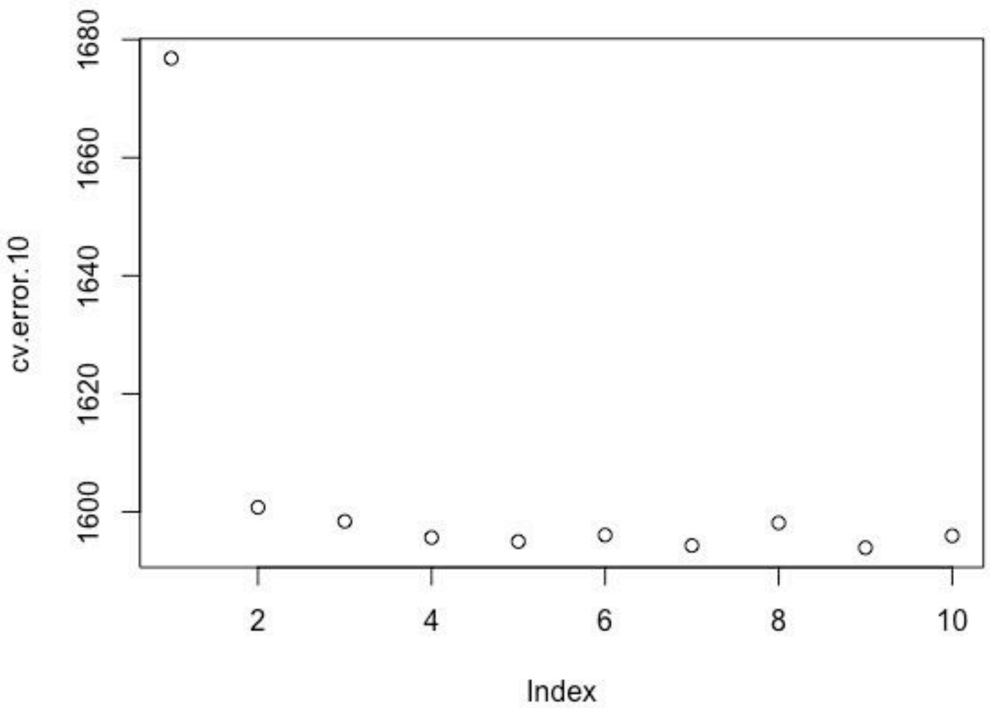
3.b.

```
boot.fn = function(data,index){
  return(coef(glm(default~income+balance, data=data, family=binomial, subset=index)))
}
```

3.c. The estimated standard errors for the coefficients of income and balance from 1000 bootstrap samples are 4.817789*10-6 and 2.341719*10-4 respectively.

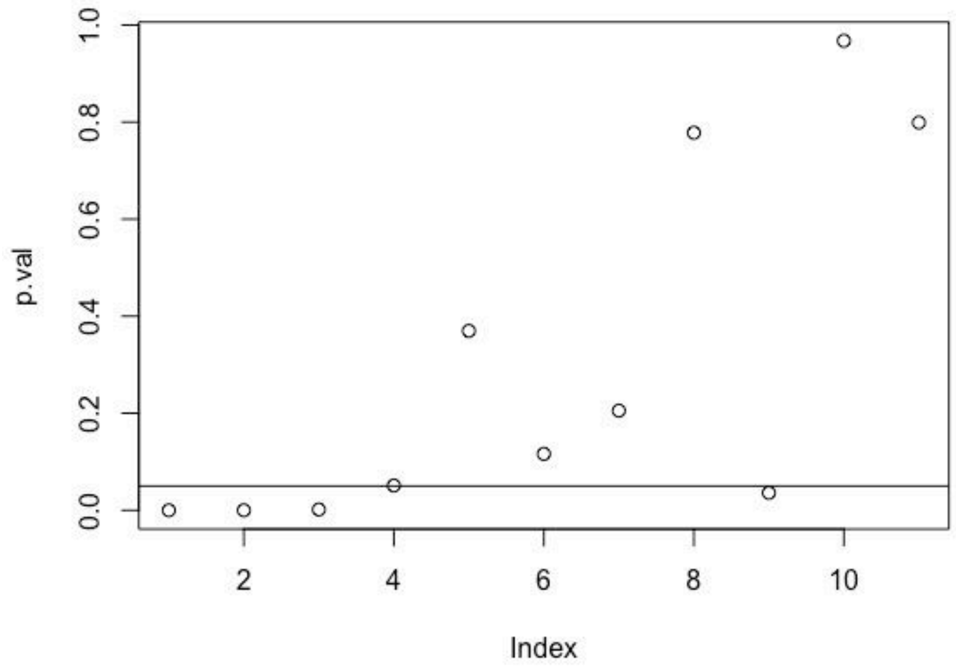
3.d. The estimates from the glm and bootstrap for the coefficients of income and balance are very similar. This is due to the fact that a large number of bootstrap samples were used to compute the estimates.

4.a.
Below is a plot of the degree of polynomial against its 10-fold cross validation error.



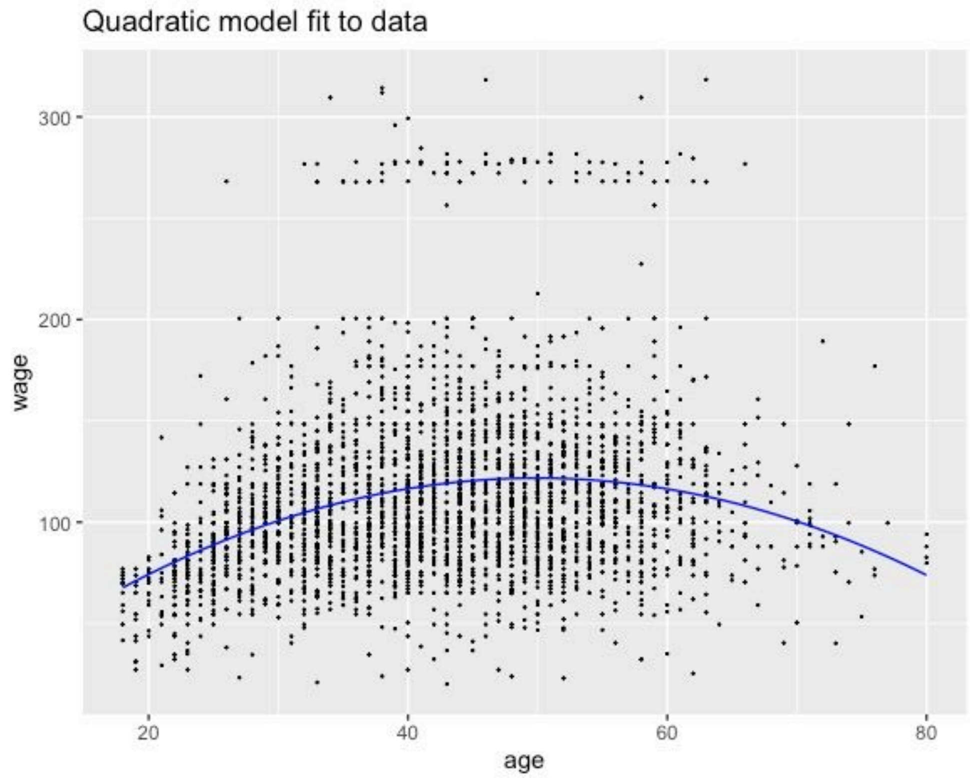
From the above plot it is clear to see that there is a large decrease in the cross-validation error by choosing a quadratic model instead of a linear model however there is minimal decrease in the error after this. Therefore, I have chosen a quadratic model.

Below is a plot of the degree of the polynomial against the p-value from the anova test:



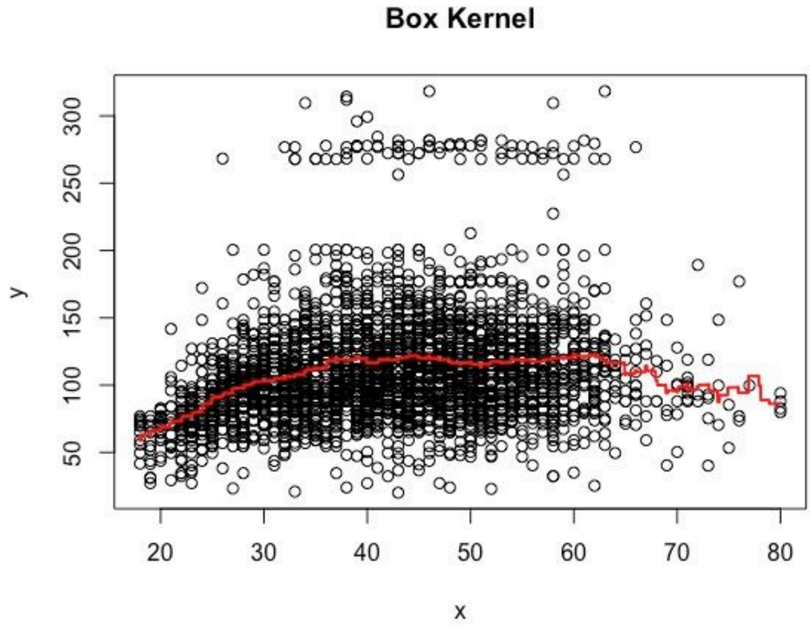
The horizontal line in the above plot is 0.05. Using the above results from anova the best choice of degree of polynomial would be a cubic model.

The plot of the quadratic model is shown below:

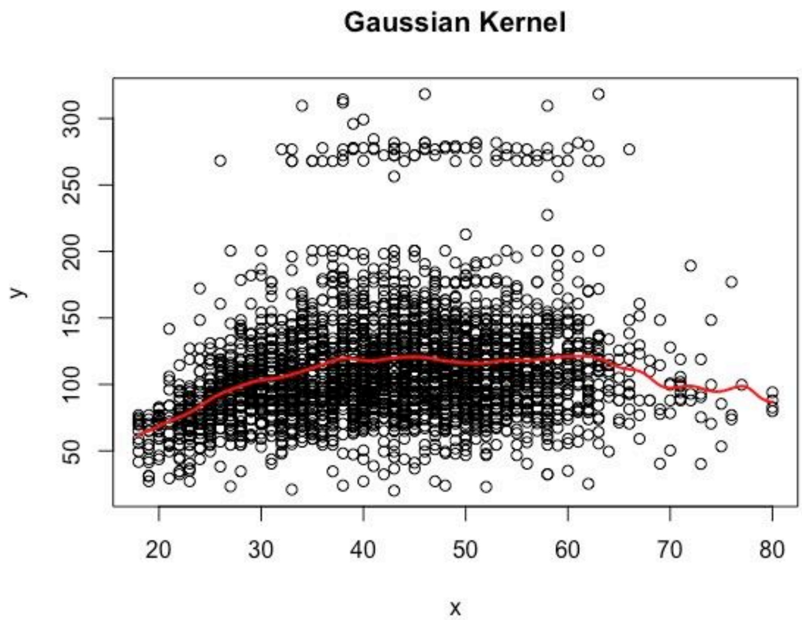


4.b. Using the hcv function within the sm package in R the optimal bandwidth calculated was h := 3.823211. I used this optimal value of bandwidth with the ksmooth function to fit a locally constant kernel regression estimator using two separate kernels (in both of the plots below x=Age and y=Wage).

<u>Box Kernel:</u>



<u>Gaussian Kernel:</u>



From the above plots it is clear to see that the box kernel is overfitted especially at the right boundary where there are less data points, compared to the gaussian kernel which is a much smoother fit overall. However, the gaussian kernel fit still suffers from overfitting at the right boundary.