Ans1) we need to calculate the MLE estimates for

$\pi, \theta_y^{(1)}, \theta_y^{(2)}$

by solving

$$f = \underset{\pi\, \theta_y^{(1)}\, \theta_y^{(2)}}{\arg\max} \sum_{i=1}^{n} \log\left(\pi^{y_i}(1-\pi)^{1-y_i}\right) + \sum_{i=1}^{n} \log\left[\left(\theta_{y_i}^{(1)}\right)^{x_{i1}} \left(1-\theta_{y_i}^{(1)}\right)^{1-x_{i1}}\right]$$

$$+ \sum_{i=1}^{n} \log\left[\theta_{y_i}^{(2)} \left(x_{i2}\right)^{-(\theta_{y_i}^{(2)}+1)}\right]$$

(a) Differentiating wrt $\pi$ and setting the derivate to zero we get

$$\frac{\Sigma y_i}{\pi} + \frac{\Sigma (1-y_i)(-1)}{\pi} = 0$$

$$\boxed{\hat{\pi} = \Sigma y_i/n}$$

(b) $$\sum_{i=1}^{n} \log\left[\left(\theta_{y_i}^{(1)}\right)^{x_{i1}} \left(1-\theta_{y_i}^{(1)}\right)^{1-x_{i1}}\right]$$

$$= \sum_{i=1}^{n} x_{i1} \log \theta_{y_i}^{(1)} + \sum_{i=1}^{n} (1-x_{i1}) \log\left(1-\theta_{y_i}^{(1)}\right)$$

$$= \sum_{i=1}^{n} x_{i1} \log\left(\theta_y^{(1)}\right) I(y_i=y) + \sum_{i=1}^{n} (1-x_{i1}) \log\left(1-\theta_y^{(1)}\right) I(y_i=y)$$

Differentiating f with respect to $\theta_y^{(1)}$ and setting the derivative to zero we get

$$= \frac{\sum_{i=1}^{n} x_{i1}\, I(y_i=y)}{\theta_y^{(1)}} + \frac{\sum_{i=1}^{n} (1-x_{i1})\, I(y_i=y)}{\left(1-\theta_y^{(1)}\right)}(-1) = 0$$

$$\boxed{\hat{\theta}_y^{(1)} = \frac{\Sigma x_{i1}\, I(y_i=y)}{\Sigma I(y_i=y)}}$$

(c)

Now $\sum_{i=1}^{n} \log\left[\theta_{y_i}^{(2)} \, (x_{i2})^{-\theta_{y_i}^{(2)}+1}\right]$

$$\Rightarrow \sum \log \theta_y^{(2)} I(y_i=y) - \sum(\theta_y^{(2)}+1) \log x_{i2} \, I(y_i=y)$$

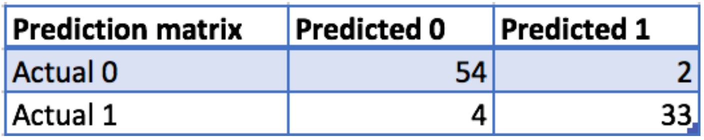
Now differentiating wrt $\theta_y^{(2)}$ and setting the derivative to zero

$$\frac{\sum I(y_i=y)}{\theta_y^{(2)}} = \sum \log(x_{i2}) \, I(y_i=y)$$

$$\boxed{\hat{\theta}_y^{(2)} = \frac{\sum I(y_i=y)}{\sum \log(x_{i2}) \, I(y_i=y)}}$$

**Q2)**

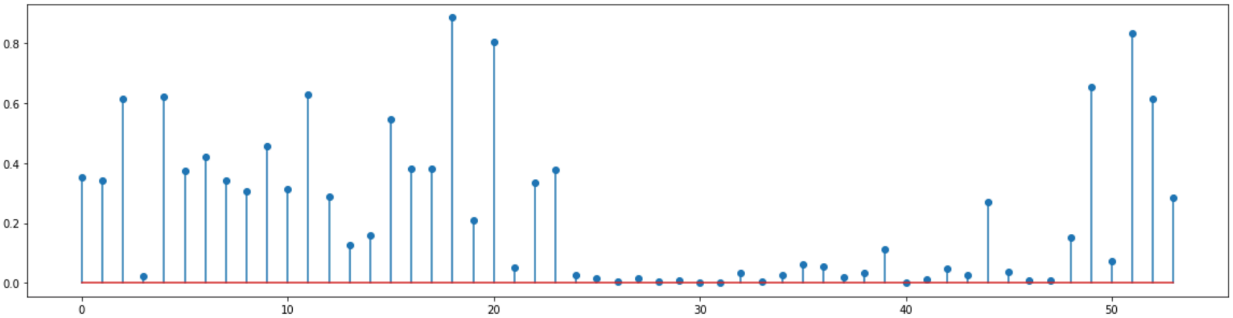
**a)**



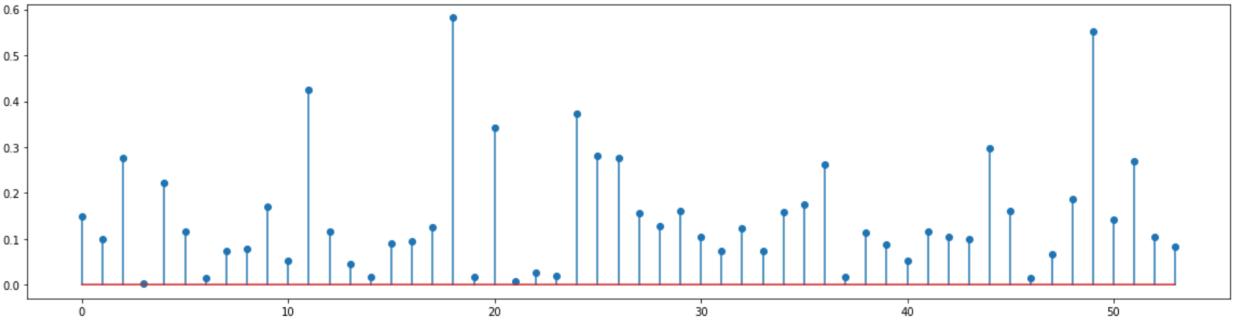
| Prediction matrix | Predicted 0 | Predicted 1 |
|---|---|---|
| Actual 0 | 54 | 2 |
| Actual 1 | 4 | 33 |

The prediction accuracy is 93.54%.

**b)** Stem plot for class 1

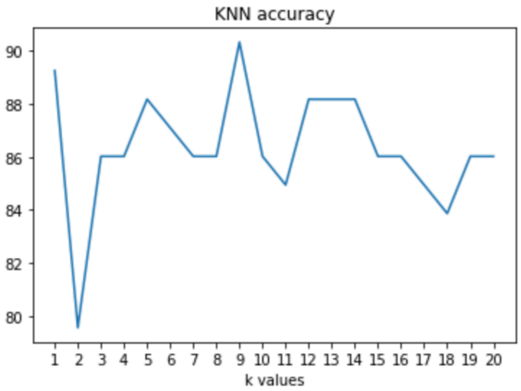


Stem plot for class 0

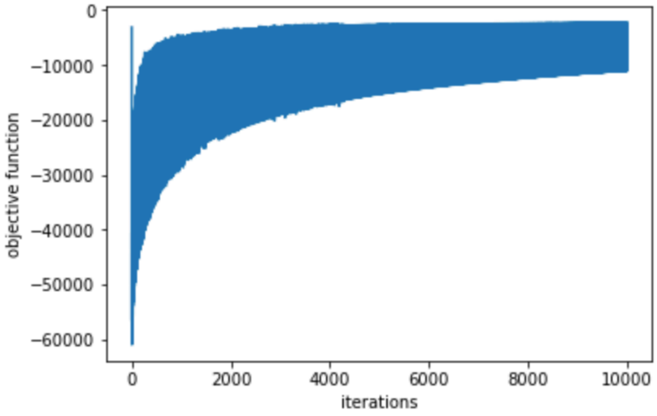


The value for the theta parameters for dimension 16 and 52 indicate that spam mails are more likely to contain " ! " and word "free".

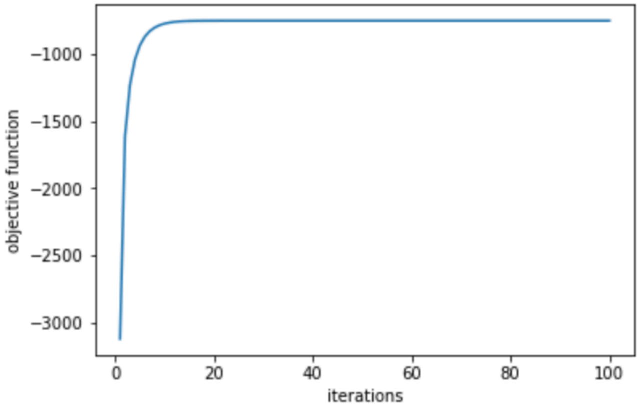
**c)** KNN accuracy plot



**d)**



**e)**



The accuracy for the newton method on test data is 91.39%.