(a) $\hat{\pi} = \underset{\hat{\pi}}{\operatorname{argmax}} \sum_{i=1}^{n} \ln\left(\pi^{y_i} (1-\pi)^{1-y_i} \right)$

$= \underset{\hat{\pi}}{\operatorname{argmax}} \sum_{i=1}^{n} \left(y_i \ln \pi + (1-y_i) \ln(1-\pi) \right) I(\pi_i = \pi)$

$\underset{\hat{\pi}}{\operatorname{argmax}} \sum_{i=1}^{n} y_i \ln(\pi) I(\pi_i = \pi) + \ln(1-\pi) I(\pi_i = \pi) - y_i \ln(1-\pi) I(\pi_i = \pi)$

$\Rightarrow \frac{1}{\hat{\pi}} \sum y_i I(\pi_i = \pi) - \left(\sum I(\pi_i = \pi) - \sum y_i I(\pi_i = \pi) \right) \frac{1}{1-\hat{\pi}} = 0$

$(1-\hat{\pi}) \sum y_i I(\pi_i = \pi) - \hat{\pi} \left(\sum I(\pi_i = \pi) - \sum y_i I(\pi_i = \pi) \right) = 0$

$\sum y_i I(\pi_i = \pi) - \hat{\pi} \sum \cancel{y_i} I(\pi_i = \pi) - \hat{\pi} \sum I(\pi_i = \pi) + \hat{\pi} \sum \cancel{y_i} I(\pi_i = \pi) =$

$\Rightarrow \sum y_i I(\pi_i = \pi) - \hat{\pi} \sum I(\pi_i = \pi) = 0$

$\hat{\pi} = \frac{\sum y_i I(\pi_i = \pi)}{\sum I(\pi_i = \pi)}$

(b) $\hat{\theta}_y^{(1)} = \arg\max_{\hat{\theta}_y^{(1)}} \sum_{i=1}^{n} \ln\left(p(x_{i1} \mid \theta_{y_i}^{(1)})\right)$

$= \arg\max_{\hat{\theta}_y^{(1)}} \sum_{i=1}^{n} \left(\ln\left[(\theta_{y_i}^{(1)})^{x_{i1}} (1-\theta_{y_i}^{(1)})^{1-x_{i1}}\right]\right) I(y_i = y)$

$= \arg\max_{\hat{\theta}_y^{(1)}} \sum_{i=1}^{n} x_{i1}\, I(y_i = y) \ln(\theta_{y_i}^{(1)}) + (1-x_{i1}) \ln(1-\theta_y^{(1)})\, I(y_i = y)$

$\frac{1}{\hat{\theta}_{y_i}^{(1)}} \sum x_{i1}\, I(y_i = y) - \left(\sum I(y_i = y) - \sum x_{i1}\, I(y_i = y)\right) \frac{1}{1-\hat{\theta}_y^{(1)}} = 0$

$$\hat{\theta}_y^{(1)} = \frac{\sum_{i=1}^{n} x \cdot I(y_i = y)}{\sum_{i=1}^{n} I(y_i = y)}$$

(c) $\hat{\theta}_y^{(2)} = \arg\max_{\theta_y^{(2)}} \sum_{i=1}^{n} \ln\left(p(x_{i2} \mid \theta_{y_i}^{(2)})\right)$

$= \arg\max_{\hat{\theta}_y^{(2)}} \sum_{i=1}^{n} \left(\ln\left[\theta_{y_i}^{(2)} (x_{i2})^{-(\theta_{y_i}^{(2)}+1)}\right]\right) I(y_i = y)$

$= \arg\max_{\theta_y^{(2)}} \sum_{i=1}^{n} \left(-(\theta_{y_i}^{(2)}+1) \ln(x_{i2})\, I(y_i = y) + \ln(\theta_y^{(2)})\, I(y_i = y)\right)$

$= \arg\max_{\theta_y^{(2)}} \ln(\theta_y^{(2)}) \sum_{i=1}^{n} I(y = y_i) - \theta_y^{(2)} \sum \ln(x_{i2})\, I(y_i = y) - \sum \ln(x_{i2})\, I(y_i = y)$

$= \frac{1}{\hat{\theta}_y^{(2)}} \sum_{i=1}^{n} I(y = y_i) - \sum \ln(x_{i2})\, I(y_i = y) = 0$

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

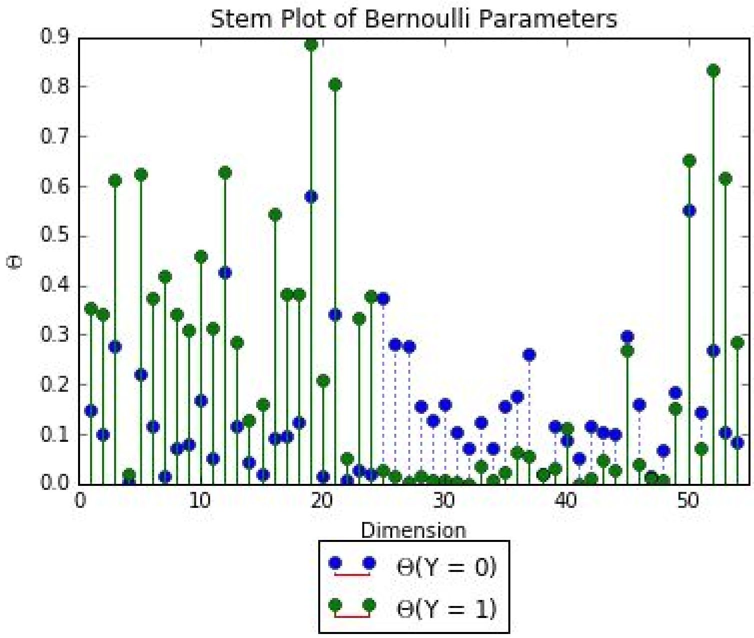
Q2)

a) Naive _ Bayes - Confusion Matrix

	0	1
0	54	2
1	5	32

Prediction Accuracy = 97.47%

b) Stem Plot :



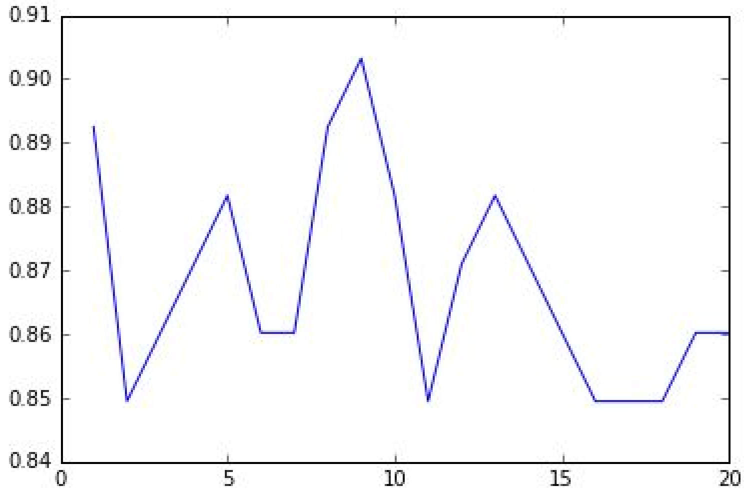
Parameter : 16 is the word "Free" and Parameter 52 is the character '!' .

The theta_not_spam Bernoulli value for "Free" is : 0.0911
The theta_spam Bernoulli value for "Free" is : 0.5450

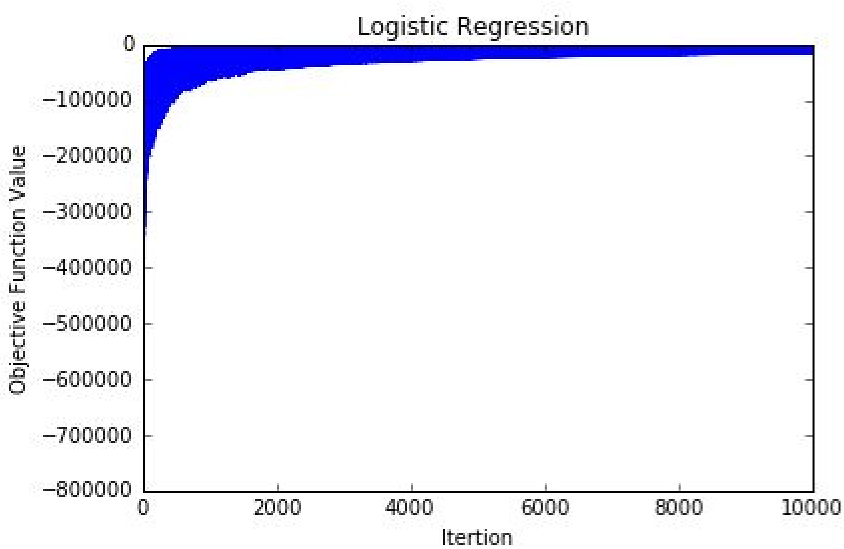
The theta_not_spam value for "!" is : 0.2690
The theta_spam value for "!" is : 0.8333

Based on the Stem plot , the difference in Thetas for these two parameters is the highest among all the other parameters.
Since the Theta_spam is higher for both the variables it can be concluded that an email with the word Free or '!' are **<u>more likely to be classified as Spam</u>** than any other variable.

c) KNN - Accuracy Plot with for N = 1 to 20 .



d) Logistic Regression : Logistic regression objective training function L per iteration



e) Newton Method :