Problem 1

(a)

$$\mathcal{L} = \sum_{i=1}^{n} \ln p(y_i \,|\pi) + \sum_{i=1}^{n} \ln p(x_{i1} \,|\theta_{y_i}^{(1)}) + \sum_{i=1}^{n} \ln p(x_{i2} \,|\theta_{y_i}^{(2)})$$

$$\frac{\partial \mathcal{L}}{\partial \pi} = \frac{\partial}{\partial \pi} \sum_{i=1}^{n} \ln(\pi^{y_i} \,(1-\pi)^{1-y_i})$$

$$= \frac{\partial}{\partial \pi} \sum_{i=1}^{n} [\mathrm{y_i} \ln \pi - (1-y_i) \ln(1-\pi)]$$

$$= \sum_{i=1}^{n} \frac{y_i}{\pi} - \sum_{i=1}^{n} \frac{1-y_i}{1-\pi}$$

$At \ \dfrac{\partial \mathcal{L}}{\partial \pi} = 0$:

$$(1-\hat{\pi}) \sum_{i=1}^{n} y_i = \hat{\pi} \sum_{i=1}^{n} (1-y_i)$$

$$\sum_{i=1}^{n} y_i - \hat{\pi} \sum_{i=1}^{n} y_i = n\hat{\pi} - \hat{\pi} \sum_{i=1}^{n} y_i$$

$$\Rightarrow \boldsymbol{\hat{\pi}} = \frac{\boldsymbol{\sum_{i=1}^{n} y_i}}{\boldsymbol{n}}$$

(b)

$$\mathcal{L} = \sum_{i=1}^{n} \ln p(y_i \mid \pi) + \sum_{i=1}^{n} \ln p(x_{i1} \mid \theta_{y_i}^{(1)}) + \sum_{i=1}^{n} \ln p(x_{i2} \mid \theta_{y_i}^{(2)})$$

$$\frac{\partial \mathcal{L}}{\partial \theta_{y_i}^{(1)}} = \frac{\partial}{\partial \pi} \sum_{i=1}^{n} \ln({\theta_y^{(1)}}^{x_{i1}} \ \left(1 - \theta_y^{(1)}\right)^{1-x_{i1}})$$

$$= \frac{\partial}{\partial \pi} \sum_{i=1}^{n} \left[x_{i1} \ln \theta_y^{(1)} + (1 - x_{i1}) \ln\left(1 - \theta_y^{(1)}\right)\right]$$

$$= \sum_{i=1}^{n} \frac{x_{i1}}{\theta_y^{(1)}} - \sum_{i=1}^{n} \frac{1 - x_{i1}}{1 - \theta_y^{(1)}}$$

$$At \ \frac{\partial \mathcal{L}}{\partial \theta_y^{(1)}} = 0:$$

$$\left(1 - \widehat{\theta_y^{(1)}}\right) \sum_{i=1}^{n} x_{i1} = \widehat{\theta_y^{(1)}} \sum_{i=1}^{n} (1 - x_{i1})$$

$$\sum_{i=1}^{n} x_{i1} - \widehat{\theta_y^{(1)}} \sum_{i=1}^{n} x_{i1} = n\widehat{\theta_y^{(1)}} - \widehat{\theta_y^{(1)}} \sum_{i=1}^{n} x_{i1}$$

$$\Rightarrow \widehat{\theta_y^{(1)}} = \boldsymbol{\frac{\sum_{i=1}^{n} x_{i1}}{n}}$$

(c)

$$\mathcal{L} = \sum_{i=1}^{n} \ln p(y_i \,|\pi) + \sum_{i=1}^{n} \ln p(x_{i1} \,|\theta_{y_i}^{(1)}) + \sum_{i=1}^{n} \ln p(x_{i2} \,|\theta_{y_i}^{(2)})$$

$$\frac{\partial \mathcal{L}}{\partial \theta_y^{(2)}} = \frac{\partial}{\partial \theta_y^{(2)}} \sum_{i=1}^{n} \ln \theta_y^{(2)} \, (x_{i2})^{-\left(\theta_y^{(2)}+1\right)}$$

$$= \frac{\partial}{\partial \theta_y^{(2)}} \sum_{i=1}^{n} \ln \theta_y^{(2)} - \left(\theta_y^{(2)} + 1\right) \ln(x_{i2})$$

$$= \frac{n}{\theta_y^{(2)}} - \sum_{i=1}^{n} \ln(x_{i2})$$

$At \ \dfrac{\partial \mathcal{L}}{\partial \theta_y^{(2)}} = 0$:

$$\boldsymbol{\widehat{\theta_y^{(2)}}} = \frac{\boldsymbol{n}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \mathbf{ln}(\boldsymbol{x_{i2}})}$$

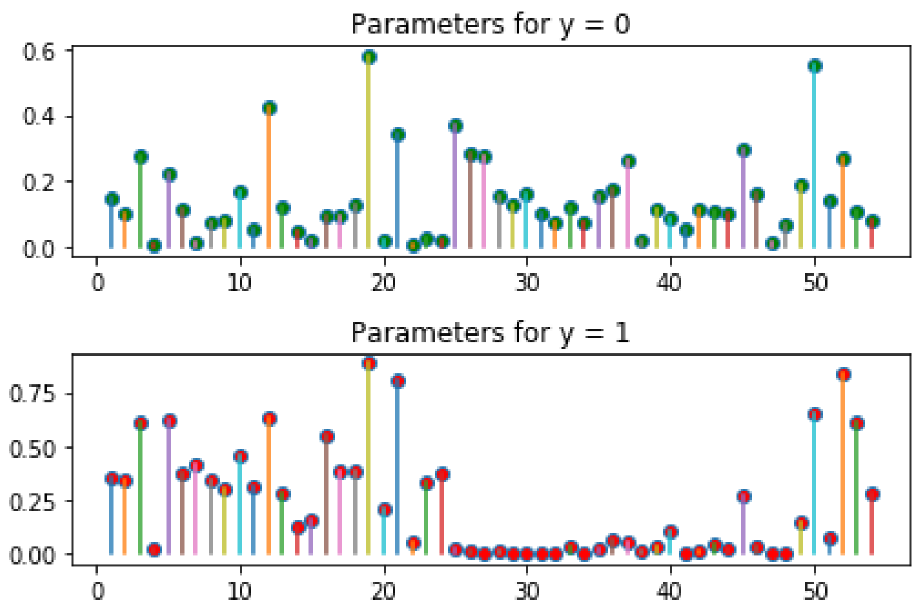
Problem 2

(a)

	Not Spam	Spam
Predicted Not spam	53	4
Predicted Spam	3	33
		Accuracy: 0.924731182796

The naïve Bayes classifier algorithm achieved a prediction accuracy of **92.47%** on the testing data.

(b)



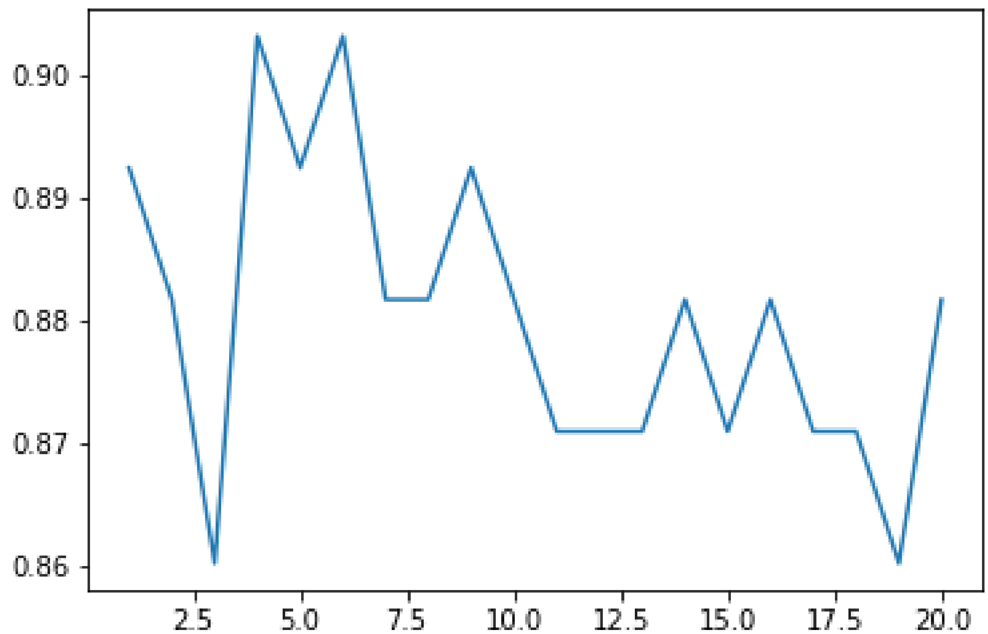
The 16th dimension corresponds to the word 'free' which has a probability of 0.0911 of appearing in a non-spam email and a probability of 0.545 of appearing in a spam email.

The 52nd dimension corresponds to the character '!' which has a probability of 0.269 of appearing in a non-spam email and a probability of 0.833 of appearing in a spam email.

These observations seem to imply that spam emails are more likely to make offers of free products and/or services in an excited tone.

(c)

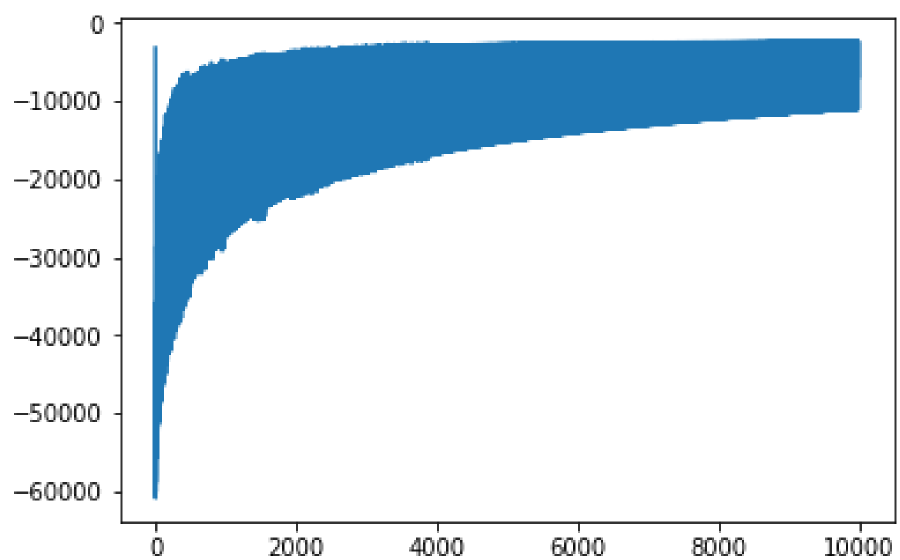
Graph of Prediction Accuracy against k



There does not seem to be any discernible pattern in the prediction accuracy, but performance seems to be best when k = 4 (accuracy = 0.903) and worst when k = 3 (accuracy = 0.86).

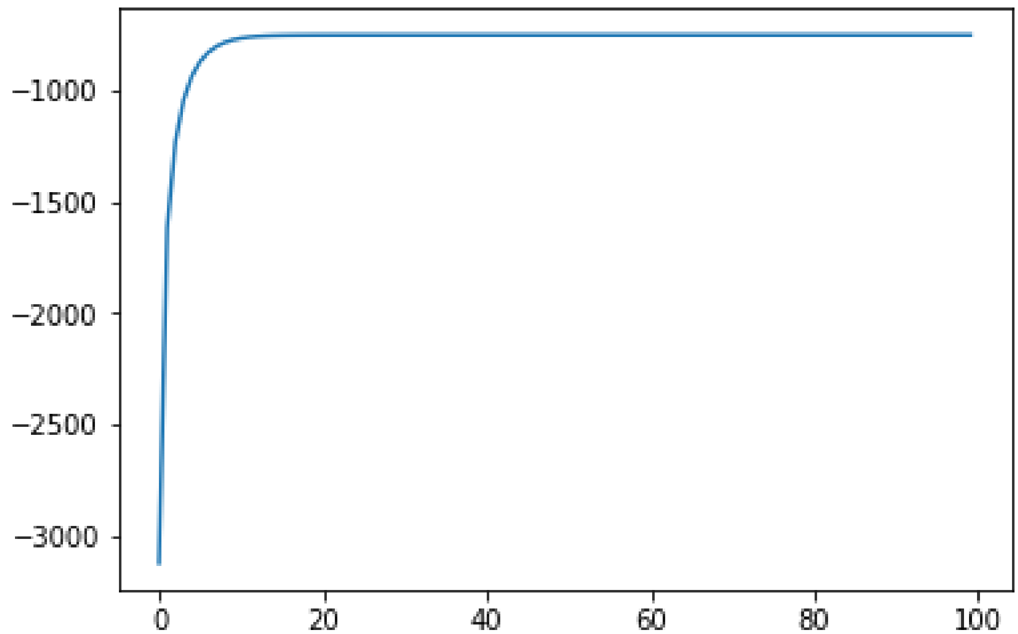
(d)

The graph below shows the value of the logistic regression objective training function $\mathcal{L}$ over 10,000 iterations of the steepest ascent algorithm.



(e)

The graph below shows the value of the logistic regression objective training function $\mathcal{L}$ over 10,000 iterations of the Newton's method.



The prediction accuracy on the testing data is 0.91397849462365588 or **91.4%.**