Arsalaan Ansari (aaa2325)

1 (a)

The portion of the objective is:

$$LL_i = \sum_{i=1}^{n} \ln (p(y_i \mid \tilde{\pi})$$

$$\nabla_{\tilde{\pi}} \sum_{i=1}^{n} \ln \left[\tilde{\pi}^{y_i} (1-\tilde{\pi})^{1-y_i} \right] = 0$$

$$\nabla_{\tilde{\pi}} \sum_{i=1}^{n} y_i \ln \tilde{\pi} + (1-y_i) \ln (1-\tilde{\pi}) = 0$$

$$\sum_{i=1}^{n} \frac{y_i}{\tilde{\pi}} - \frac{1}{1-\tilde{\pi}} + \frac{y_i}{1-\tilde{\pi}} = 0$$

$$\sum_{i=1}^{n} \frac{y_i (1-\tilde{\pi})}{\tilde{\pi}(1-\tilde{\pi})} + \frac{(y-1)(\tilde{\pi})}{\tilde{\pi}(1-\tilde{\pi})} = 0$$

$$\sum_{i=1}^{n} y_i - \cancel{y_i \tilde{\pi}} - \tilde{\pi} + \cancel{y_i \tilde{\pi}} = 0$$

$$\sum_{i=1}^{n} (y_i - \tilde{\pi}) = 0$$

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

class prior for $y = 1$ is $\tilde{\pi}$

class prior for $y = 0$ is $1 - \tilde{\pi}$

1(b)

The portion of the objective is

$$LL_2 = \sum_{i=1}^{n} \ln p(x_{i1} \mid \theta_{y_i}^{(1)})$$

This is a bernoulli distribution (like the one we derived before) therefore the $\theta_y^{(1)}$ parameter for each class is the mean of x_{i1} values for that class:

$$\theta_{y=0}^{(1)} = \frac{\sum_{i=1}^{n} x_{i1} \, 1(y_i = 0)}{\sum_{i=1}^{n} 1(y_i = 0)}$$

$1(\cdot)$ is the indicator function.

and

$$\theta_{y=1}^{(1)} = \frac{\sum_{i=1}^{n} x_{i1} \, 1(y_i = 1)}{\sum_{i=1}^{n} 1(y_i = 1)}$$

Therefore if we leave y arbitrary

$$\boxed{\theta_y^{(1)} = \frac{\sum_{i=1}^{n} x_{i1} \, 1(y_i = y)}{\sum_{i=1}^{n} 1(y_i = y)}}$$

where $y = \{0,1\}$

1(c)

The relevant portion of the objective is:

$$LL_3 = \sum_{i=1}^{n_y} \ln p(x_{i2} \mid \theta_{y_i}^{(2)})$$ where n_y is # of observations for class $y = \{0,1\}$

First solving for the MLE

$$\nabla_{\theta_{y_i}^{(2)}} \sum_{i=1}^{n_y} \ln \left[(\theta_{y_i}^{(2)}) \cdot (x_{i2})^{-(\theta_{y_i}^{(2)}+1)} \right] = 0$$

$$\nabla_{\theta_{y_i}^{(2)}} \sum_{i=1}^{n_y} \ln \theta_{y_i}^{(2)} + (-(\theta_{y_i}^{(2)}+1)) \ln(x_{i2}) = 0$$

$$\sum_{i=1}^{n_y} \frac{1}{\theta_{y_i}^{(2)}} - \ln(x_{i2}) = 0$$

$$\sum_{i=1}^{n_y} \frac{1}{\theta_{y_i}^{(2)}} = \sum_{i=1}^{n_y} \ln(x_{i2})$$

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

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

Therefore we can find the class conditionals

$$\theta_{y=0}^{(2)} = \frac{\sum_{i=1}^{n} \ln(x_{i2}) \mathbb{1}(y_i=0)}{\sum_{i=1}^{n} \mathbb{1}(y_i=0)}, \quad \theta_{y=1}^{(2)} = \frac{\sum_{i=1}^{n} \ln(x_{i2}) \mathbb{1}(y_i=1)}{\sum_{i=1}^{n} \mathbb{1}(y_i=1)}$$

For an arbitrary y:

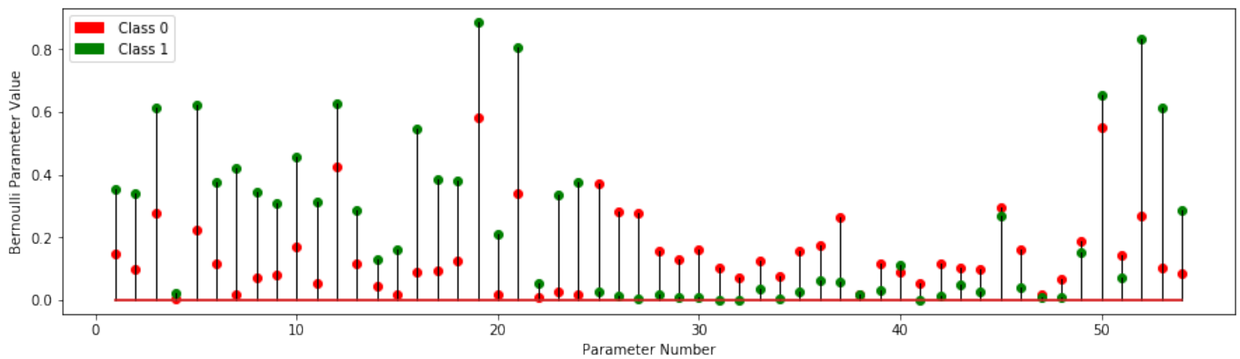
$$\boxed{\theta_y^{(2)} = \frac{\sum_{i=1}^{n} \ln(x_{i2}) \mathbb{1}(y_i=y)}{\sum_{i=1}^{n} \mathbb{1}(y_i=y)}}$$

2 (a): Confusion Matrix

Class	Predicted 0	Predicted 1
Actual 0	54	2
Actual 1	5	32

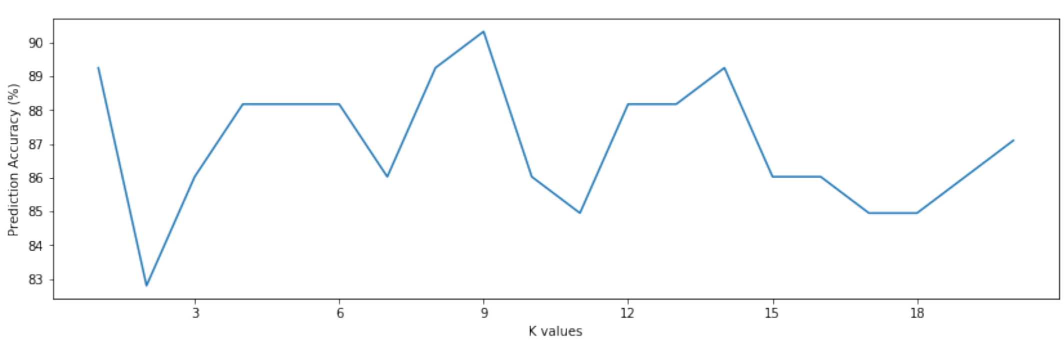
Prediction Accuracy: 92.473

2 (b): Stem Plot



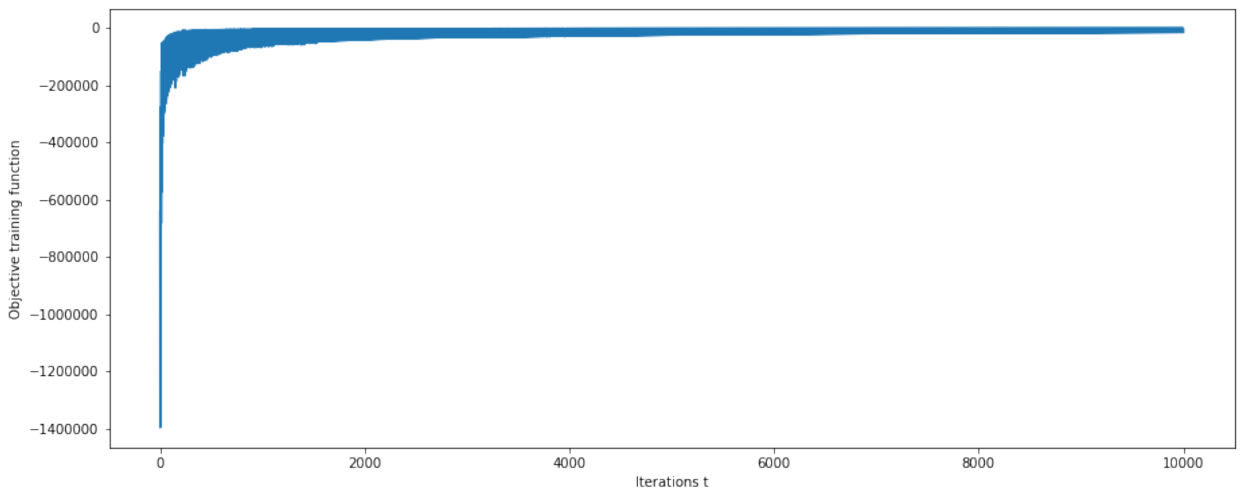
Looking at features 16 and 52, they represent the word ‘free’ and ‘!’ respectively. In both of these features the Bernoulli parameter for class 1 (Spam) represented by green dots on the stem plot are considerably larger than the Bernoulli parameter for class 0 (Not Spam). This indicates the word ‘free’ and the character ‘!’ occur much more frequently in spam emails than non-span emails. For instance the plot shows that the probability of seeing the word ‘free’ in spam emails is .545 compared to the probability of seeing it in non-spam emails being 0.0911. Similarly the probability of seeing the character ‘!’ in spam emails is 0.833 whereas the probability is only .269.

2 (c): KNN Plot (Prediction Accuracy vs K)



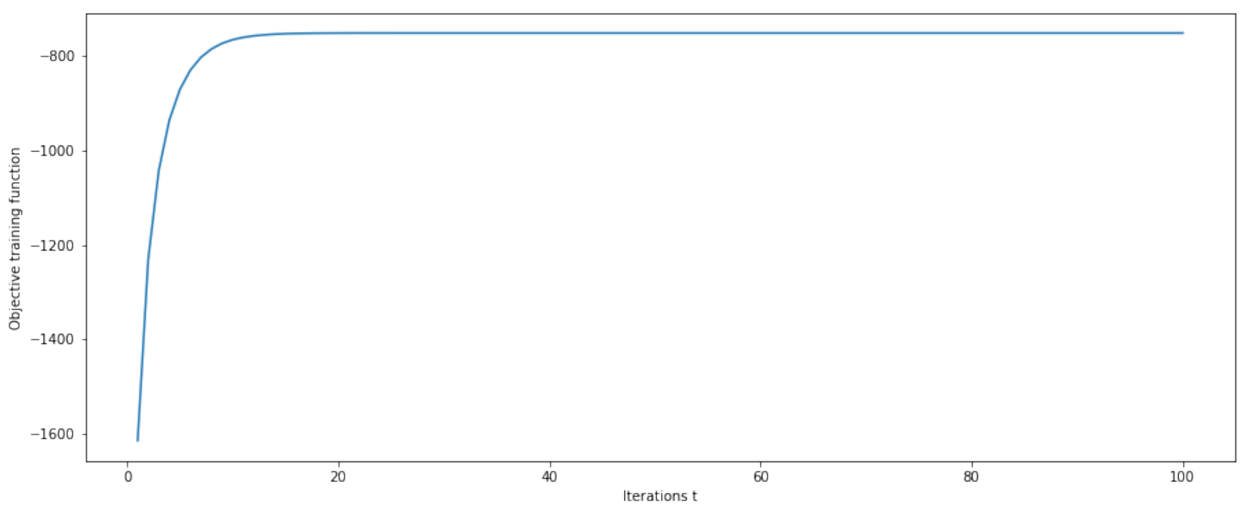
Ties were broken randomly in the KNN algorithm

2 (d): Logistic Regression Plot



Test Data Accuracy: 74.19%

2 (e): Newton Method Plot



Test Data Accuracy: 91.40%