HW2 Report

Name: Wenbo Song
UNI: ws2505

Problem 1

$$\hat{\pi}, \hat{\theta}_y^{(1)}, \hat{\theta}_y^{(2)} = argmax_{\hat{\pi}, \hat{\theta}_y^{(1)}, \hat{\theta}_y^{(2)}} \sum_{i=1}^{n} lnp(y_i|\pi) + \sum_{i=1}^{n} lnp(x_{i1}|\hat{\theta}_y^{(1)}) + \sum_{i=1}^{n} lnp(x_{i2}|\hat{\theta}_y^{(2)})$$

(a) Derive $\hat{\pi}$ using the objective above

$$\hat{\pi} = argmax_{\hat{\pi}} \sum_{i=1}^{n} lnp(y_i|\pi)$$

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

$set\ the\ gradient\ to\ 0$

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

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

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

(b) Derive $\hat{\theta}_y^{(1)}$ using the objective above. Derive this leaving y arbitrary.

$$\hat{\theta}_y^{(1)} = argmax_{\hat{\theta}_y^{(1)}} \sum_{i=1}^{n} lnp\left(x_{i1}\middle|\hat{\theta}_y^{(1)}\right)$$

$$= argmax_{\hat{\theta}_y^{(1)}} \sum_{i=1}^{n} x_{i1} ln\hat{\theta}_y^{(1)} + (1 - x_{i1})\ln\left(1 - \hat{\theta}_y^{(1)}\right)$$

$$= argmax_{\hat{\theta}_y^{(1)}} \left(\sum_{i=1}^{n} x_{i1} ln\hat{\theta}_y^{(1)} + (1 - x_{i1})\ln\left(1 - \hat{\theta}_y^{(1)}\right)\right) 1(y_i = y)$$

$set\ the\ gradient\ to\ 0$

$$\nabla_\pi \left(\sum_{i=1}^{n} x_{i1} ln\hat{\theta}_y^{(1)} + (1 - x_{i1})\ln\left(1 - \hat{\theta}_y^{(1)}\right)\right) 1(y_i = y) = 0$$

$$\hat{\theta}_y^{(1)} = \frac{\sum_{i=1}^{n} x_{i1} 1(y_i = y)}{n_y}$$

$where\ n_y = \sum_{i=1}^{n} 1(y_i = y)$

(c) Derive $\hat{\theta}_y^{(2)}$ using the objective above. Derive this leaving y arbitrary.

$$\hat{\theta}_y^{(2)} = argmax_{\hat{\theta}_y^{(2)}} \sum_{i=1}^{n} lnp\left(x_{i2} \middle| \hat{\theta}_y^{(2)}\right)$$

$$= argmax_{\hat{\theta}_y^{(2)}} \left(\sum_{i=1}^{n} ln\hat{\theta}_y^{(2)} - \left(1 - \hat{\theta}_y^{(2)}\right) lnx_{i2}\right) 1(y_i = y)$$

$set\ the\ gradient\ to\ 0$

$$\hat{\theta}_y^{(2)} = n_y \sum_{i=1}^{n} \frac{1}{lnx_{i2}} \quad 1(y_i = y)$$

$where\ n_y = \sum_{i=1}^{n} 1(y_i = y)$

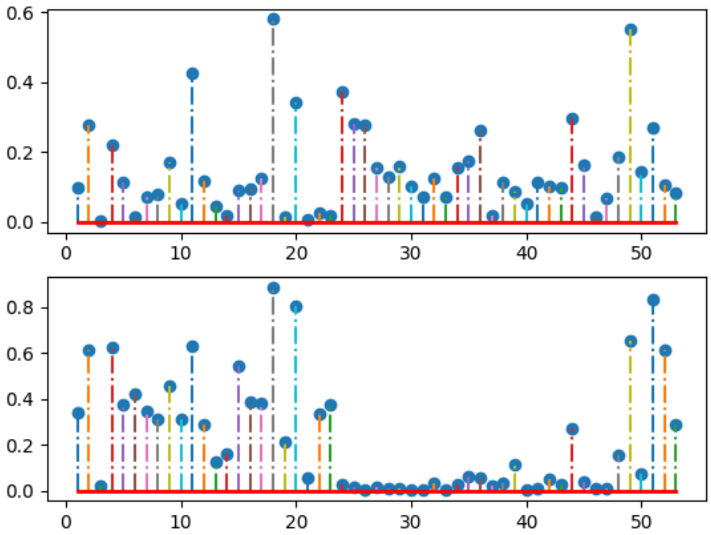
Problem 2

(a) ImplementthenaiveBayesclassifierdescribedaboveonthetrainingdataandmakepredictionso n the testing data. In a 2 × 2 table, write the number of times that you predicted a class y data point (ground truth) as a class y′ data point (model prediction) in the (y, y′)-th cell of the table, where y and y′ can be either 0 or 1. Next to your table, write the prediction accuracy

	Actual to be 0	Actual to be 1
Predict to be 0	54	2
Predict to be 1	5	32

From the table, we can know that the accuracy is $acc = \frac{54+32}{93} = 0.9247 = 92.47\%$

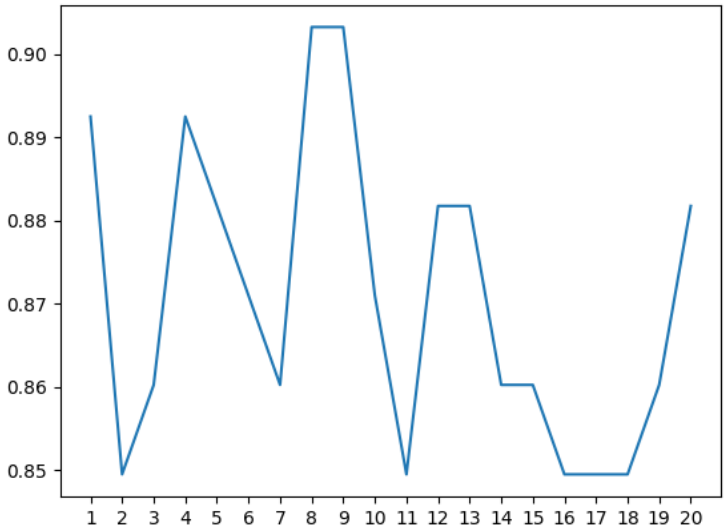
(b) In one figure, show a stem plot (stem() in Matlab) of the 54 Bernoulli parameters for each class. Use the file "spambase.names" to make an observation about dimensions 16 and 52.



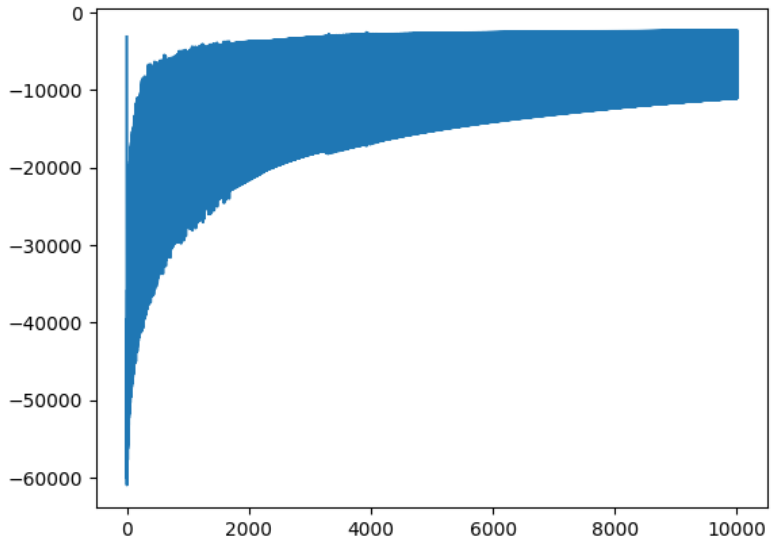
Above is for class 0 and below is for class 1, using the spambase.names we can know that dimension 16 is frequency of word “free” and dimension 52 is frequency of char “!”.
We can see from the figure that the in spam e-mail the frequency of “free” and “!” is obviously higher than that in non-spam e-mail, which indicates that the spam e-mail are more frequently use word “free” and char “!”

(c) Implement the k-NN algorithm for k = 1, . . . , 20. Use the l1 distance for this problem (sum of the absolute values of the differences). Plot the prediction accuracy as a function of k.

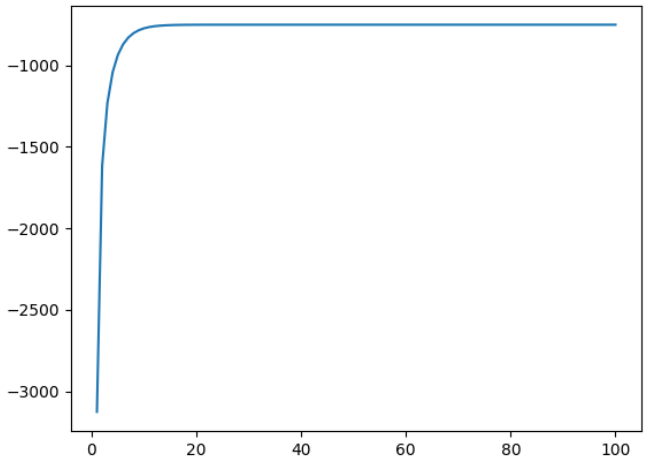
The figure is shown as below:



(d) Implement the steepest ascent algorithm given in Lecture 9. Use an iteration-dependent step size ηt. Run your algorithm for 10,000 iterations and plot the logistic regression objective training function L per iteration.
For 10000 iterations, the pattern of the objective function L is shown as below:



(e) Finally, implement a gradient method called "Newton's method". Plot the objective function L on the training data as a function t+1 of t = 1, . . . , 100. Below the plot give the prediction accuracy on the testing data.
The figure is shown as below:



And the accuracy is 91.40%