# ELEN 4903: Machine Learning
# Homework 2

Josh Rutta

February 25, 2018

**Problem 1 (written):**
**(Scan on next page)**

Josh Rutta
Prof. Paisley
ELEN 4903
HW2

$$p(y|\pi) = \pi^{y_i}(1-\pi)^{1-y_i}$$

1. $y_0 = \arg\max_y p(y_0 = y|\pi)\prod_{d=1}^{D} p_d(x_{0,d}|\theta_y^{(d)})$; $D=2$; $p_1(x_{0,1}|\theta_y^{(1)}) = (\theta_y^{(1)})^{x_{0,1}}(1-\theta_y^{(1)})^{1-x_{0,1}}$

$p_2(x_{0,2}|\theta_y^{(2)}) = \theta_y^{(2)}(x_{0,2})^{-(\theta_y^{(2)}+1)}$; $\theta_y^{(1)} \in [0,1]$; $\theta_y^{(2)} > 0$; $\pi \in [0,1]$

(a) $\hat{\pi} = \arg\max_\pi \sum_{i=1}^n \ln p(y_i|\pi) + \sum_{i=1}^n \ln p(x_{i,1}|\theta_{y_i}^{(1)}) + \sum_{i=1}^n \ln p(x_{i,2}|\theta_{y_i}^{(2)})$

$\hat{\pi} = \arg\max_\pi \sum_{i=1}^n \ln(\pi^{y_i}(1-\pi)^{1-y_i}) + \sum_{i=1}^n \ln([(\theta_{y_i}^{(1)})^{x_{i,1}}(1-\theta_{y_i}^{(1)})^{1-x_{i,1}}]) + \sum_{i=1}^n \ln(\theta_{y_i}^{(2)}(x_{i,2})^{-(\theta_{y_i}^{(2)}+1)})$

$\frac{\partial}{\partial\pi}\left(\sum_{i=1}^n[y_i\ln\pi + (1-y_i)\ln(1-\pi)] + \sum_{i=1}^n[x_{i,1}\ln\theta_{y_i}^{(1)} + (1-x_{i,1})\ln(1-\theta_{y_i}^{(1)})] + \sum_{i=1}^n[\ln(\theta_{y_i}^{(2)}) - (\theta_{y_i}^{(2)}+1)\ln(x_{i,2})]\right)$

$= \sum_{i=1}^n\left(\frac{y_i}{\pi} - \frac{(1-y_i)}{1-\pi}\right) = \frac{\sum_{i=1}^n y_i}{\pi} - \frac{n - \sum_{i=1}^n y_i}{1-\pi} = 0 \Rightarrow \frac{1-\pi}{\pi} = \frac{n-\sum y_i}{\sum y_i} \Rightarrow \frac{1}{\pi} - \cancel{1} = \frac{n}{\sum y_i} - \cancel{1} \Rightarrow \boxed{\hat{\pi} = \frac{\sum_{i=1}^n y_i}{n}}$

(b) $\hat{\theta}_{y_i}^{(1)}: \frac{\partial}{\partial\theta_{y_i}^{(1)}}\left[\sum_{i=1}^n \ln(\pi^{y_i}(1-\pi)^{1-y_i}) + \sum_{i=1}^n \ln((\theta_{y_i}^{(1)})^{x_{i,1}}(1-\theta_{y_i}^{(1)})^{1-x_{i,1}}) + \sum_{i=1}^n \ln(\theta_{y_i}^{(2)}(x_{i,2})^{-(\theta_{y_i}^{(2)}+1)})\right]$

2 cases: $y_i = 0, y_i = 1$; Let $S_0 = \{i \in \{1,\dots,n\} | y_i = 0\}$, $S_1 = \{i \in \{1,\dots,n\} | y_i = 1\}$, $n_0 = |S_0|$, $n_1 = |S_1|$

$\hat{\theta}_0^{(1)}: \frac{\partial}{\partial\theta_0^{(1)}}\left(\sum_{i=1}^n(x_{i,1}\ln\theta_{y_i}^{(1)} + (1-x_{i,1})\ln(1-\theta_{y_i}^{(1)})\right) = \sum_{i\in S_0}\left(\frac{x_{i,1}}{\theta_0^{(1)}} + \frac{(1-x_{i,1})}{1-\theta_0^{(1)}}\right) = 0$

$\Rightarrow \frac{1-\theta_0^{(1)}}{\theta_0^{(1)}} = \frac{n_0 - \sum_{i\in S_0} x_{i,1}}{\sum_{i\in S_0} x_{i,1}} \Rightarrow \frac{1}{\theta_0^{(1)}} = \frac{n_0}{\sum_{i\in S_0} x_{i,1}} \Rightarrow \boxed{\hat{\theta}_0^{(1)} = \frac{\sum_{i\in S_0} x_{i,1}}{n_0}}$

Similarly, for $\hat{\theta}_1^{(1)}: \frac{\partial}{\partial\theta_1^{(1)}}\left(\sum_{i=1}^n(x_{i,1}\ln\theta_{y_i}^{(1)} + (1-x_{i,1})\ln(1-\theta_{y_i}^{(1)})\right) = \sum_{i\in S_1}\left(\frac{x_{i,1}}{\theta_1^{(1)}} + \frac{(1-x_{i,1})}{1-\theta_1^{(1)}}\right) = 0$

$\Rightarrow \boxed{\hat{\theta}_1^{(1)} = \frac{\sum_{i\in S_1} x_{i,1}}{n_1}}$

(c) $\hat{\theta}_{y_i}^{(2)}: \frac{\partial}{\partial\theta_{y_i}^{(2)}}\left[\sum_{i=1}^n \ln(\pi^{y_i}(1-\pi)^{1-y_i}) + \sum_{i=1}^n \ln((\theta_{y_i}^{(1)})^{x_{i,1}}(1-\theta_{y_i}^{(1)})^{1-x_{i,1}}) + \sum_{i=1}^n \ln(\theta_{y_i}^{(2)}(x_{i,2})^{-(\theta_{y_i}^{(2)}+1)})\right)$

2 cases: $y_i = 0, y_i = 1$; Let $S_0 = \{i \in \{1,\dots,n\} | y_i = 0\}$, $S_1 = \{i \in \{1,\dots,n\} | y_i = 1\}$, $n_0 = |S_0|$, $n_1 = |S_1|$

$\hat{\theta}_0^{(2)}: \frac{\partial}{\partial\theta_0^{(2)}}\left(\sum_{i=1}^n(\ln\theta_{y_i}^{(2)} - (\theta_{y_i}^{(2)}+1)\ln(x_{i,2})\right) = \sum_{i\in S_0}\left(\frac{1}{\theta_0^{(2)}} - \ln(x_{i,2})\right) = \frac{n_0}{\theta_0^{(2)}} - \sum_{i\in S_0}\ln(x_{i,2}) = 0$

$\hat{\theta}_0^{(2)} = \frac{n_0}{\sum_{i\in S_0}\ln(x_{i,2})}$

$\hat{\theta}_1^{(2)}: \frac{\partial}{\partial\theta_1^{(2)}}\left(\sum_{i=1}^n(\ln\theta_{y_i}^{(2)} - (\theta_{y_i}^{(2)}+1)\ln(x_{i,2})\right) = \sum_{i\in S_1}\left(\frac{1}{\theta_1^{(2)}} - \ln(x_{i,2})\right) = \frac{n_1}{\theta_1^{(2)}} - \sum_{i\in S_1}\ln(x_{i,2}) = 0$

$\hat{\theta}_1^{(2)} = \frac{n_1}{\sum_{i\in S_1}\ln(x_{i,2})}$

**Problem 2 (coding) Figures:**

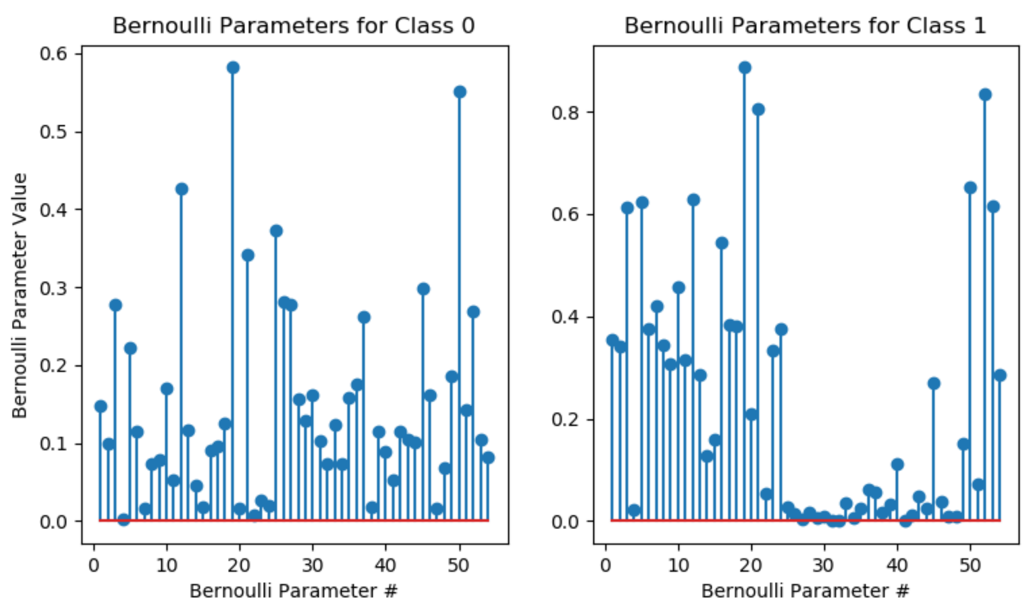
2a.

| Confusion Matrix | | |
|---|---|---|
| | Predicted 0 | Predicted 1 |
| True 0 | 54 | 2 |
| True 1 | 5 | 32 |

Prediction Accuracy:

$$86/93 = 92.46\%$$

2b.

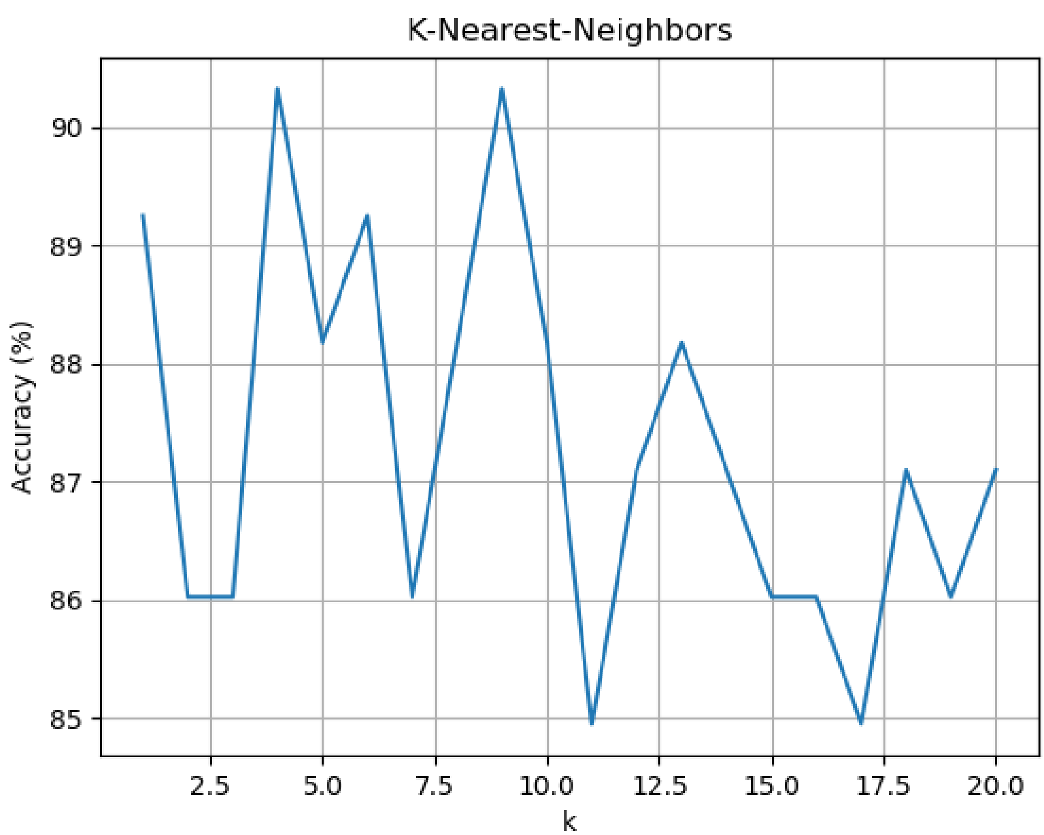


From spambase.names, we know that the Bernoulli parameters 16 and 52 represent the word ”free” and the character ”!” respectively.

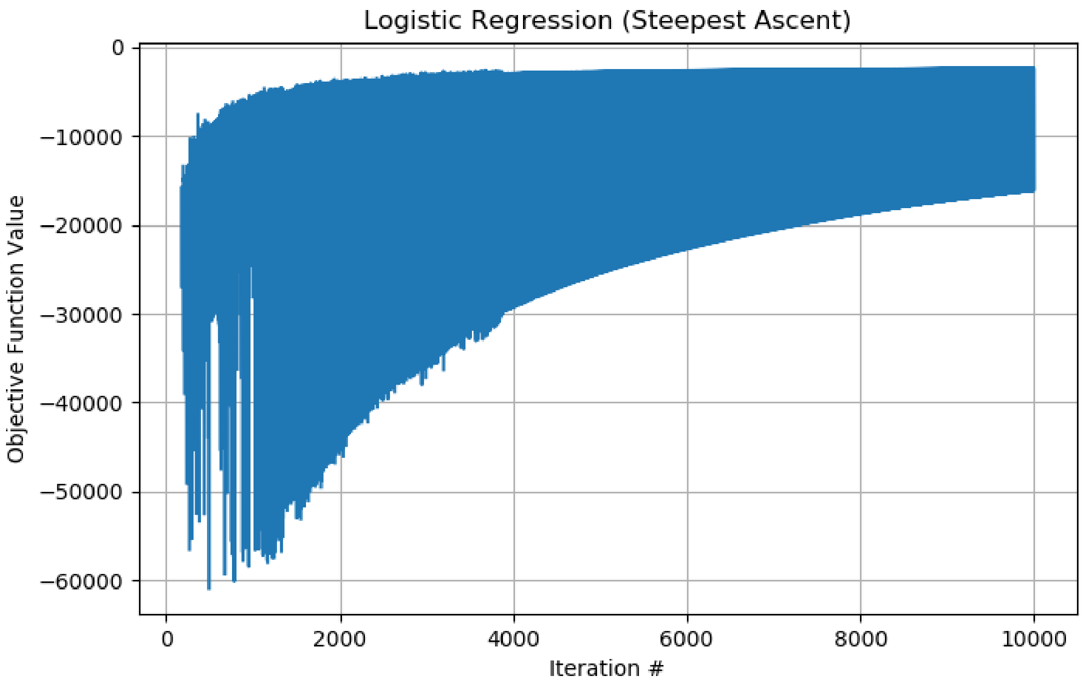
In the Class 0 Bernoulli Parameters, the word ”free” has a value of 0.091, and the character ”!” has the value 0.269. These parameters represent the probability that ”free” and ”!” are *not* in a spam email. As expected, the parameters have relatively low probabilities, meaning they'd be unlikely to show up in an email that wasn't spam.

In the Class 1 Bernoulli Parameters, the word ”free” has a value of 0.545, and the character ”!” has the value 0.833. These parameters represent the probability that ”free' and ”!” *are* in a spam email. As expected, the parameters have relatively high probabilities, meaning they are likely to show up in a spam email.

2c.

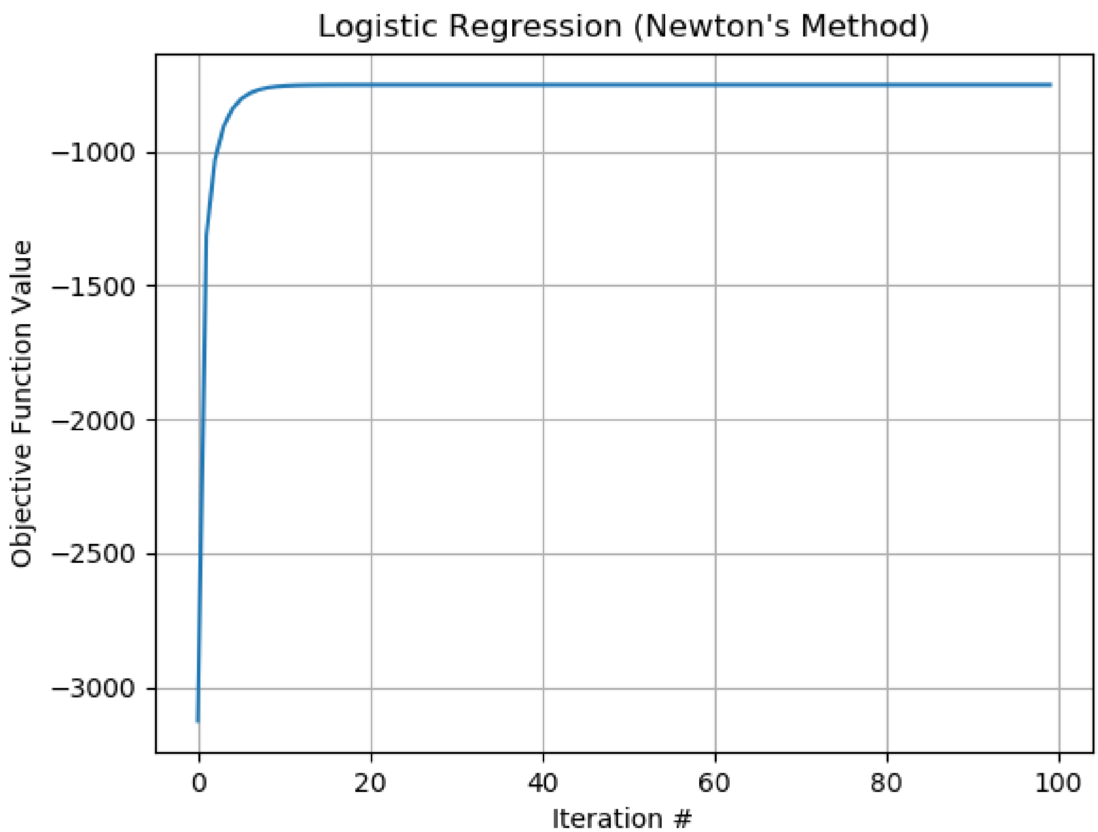


2d.



Logistic Regression (Steepest Ascent)
Objective Function Value
Iteration #
0
−10000
−20000
−30000
−40000
−50000
−60000
0
2000
4000
6000
8000
10000

2e.




Prediction Accuracy: 91.40%