HW 2 NAME: Wei Dai UNI: wd2281

1. (1) $\hat{\pi} = \arg\max_{\pi} L$.

$$\rightarrow \frac{\partial L}{\partial \pi} = \frac{\partial}{\partial \pi}\left(\sum_{i=1}^{n} \ln (P(y_i|\pi))\right)$$

$$= \frac{\partial}{\partial \pi}\left(\sum_{i=1}^{n} \ln \pi^{y_i}(1-\pi)^{1-y_i}\right)$$

$$= \frac{\partial}{\partial \pi}\left(\sum_{i=1}^{n} y_i \ln \pi + \sum_{i=1}^{n}(1-y_i)\ln(1-\pi)\right) = 0$$

$$\frac{1}{\pi}\cdot \sum_{i=1}^{n} y_i + \left(n - \sum_{i=1}^{n} y_i\right)\frac{-1}{1-\pi} = 0$$

assume

$$\rightarrow \pi = \frac{1}{n}\cdot \sum_{i=1}^{n} y_i \rightarrow \hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \left(P(Y=y|\pi) = \begin{cases} \pi & y=1 \\ 1-\pi & y=0 \end{cases} \right)$$

(2) $\theta_y^{(1)}$

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

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

$$\frac{\sum_{i=1}^{n} x_{i1}\mathbb{1}(y=y_i)}{\theta_y^{(1)}} + \frac{\sum_{i=1}^{n}(1-x_{i1})\mathbb{1}(y_i=y)}{\theta_y^{(1)}} = 0$$

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

(3) $\theta_y^{(2)}$ $\quad \frac{\partial L}{\partial \theta^{(2)}} = \frac{\partial}{\partial \theta_y^{(2)}}\left(\sum_{i=1}^{n} \ln (P(x_{i2}|\theta_{y_i}^{(2)}))\right) = \frac{\partial}{\partial \theta_y^{(2)}}\left(\sum_{i=1}^{n} \ln \theta_{y_i}^{(2)}(x_{i2})^{-(\theta_{y_i}^{(2)}+1)}\right)$

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

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

$$\rightarrow \hat{\theta}_y^{(2)} = \frac{\sum_{i=1}^{n}\mathbb{1}(y_i=y)}{\sum_{i=1}^{n}\ln x_{i2}\,\mathbb{1}(y_i=y)}$$

HW2

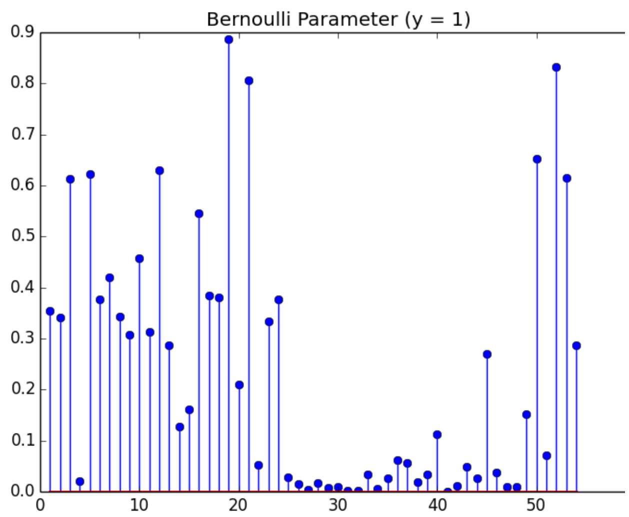
Name: Wei Dai
Uni: wd2281

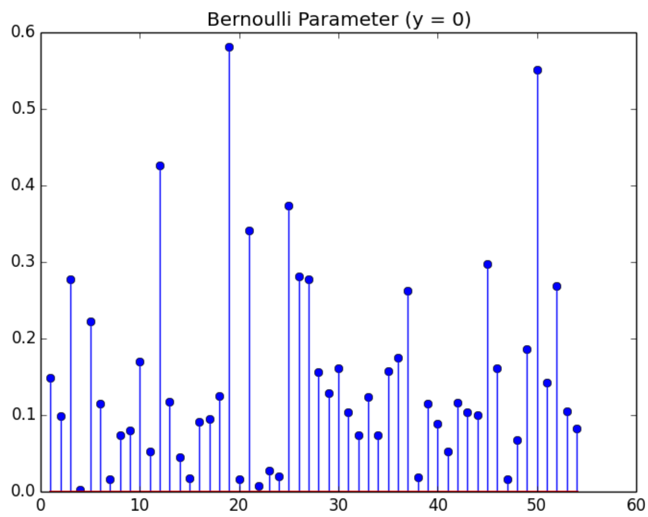
2.
a.

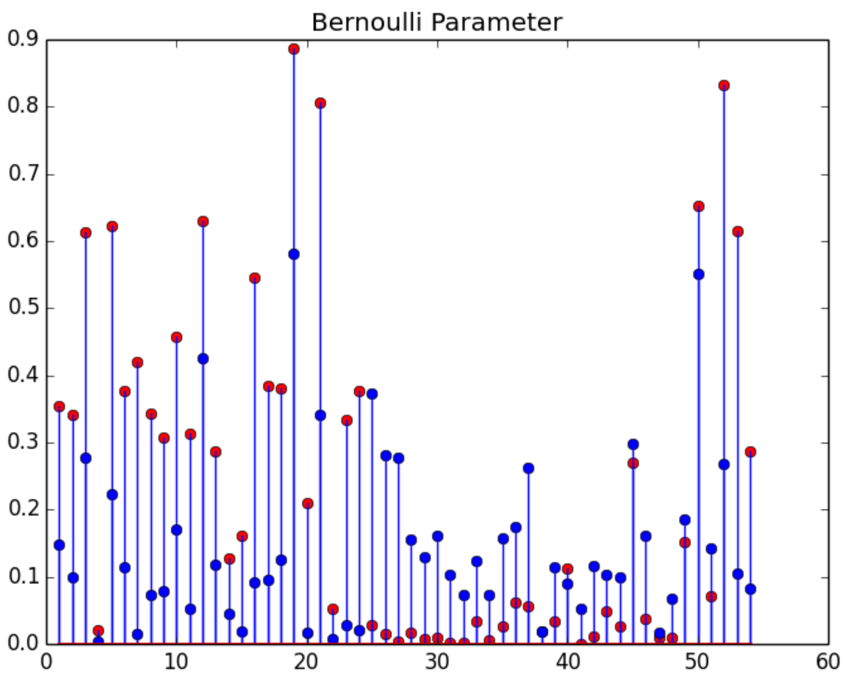
	prediction y = 0	prediction y = 1
ytest y = 0	54	2
ytest y = 1	5	32

The prediction accuracy is (54 + 32) / 93 = 0.924731182796

b.

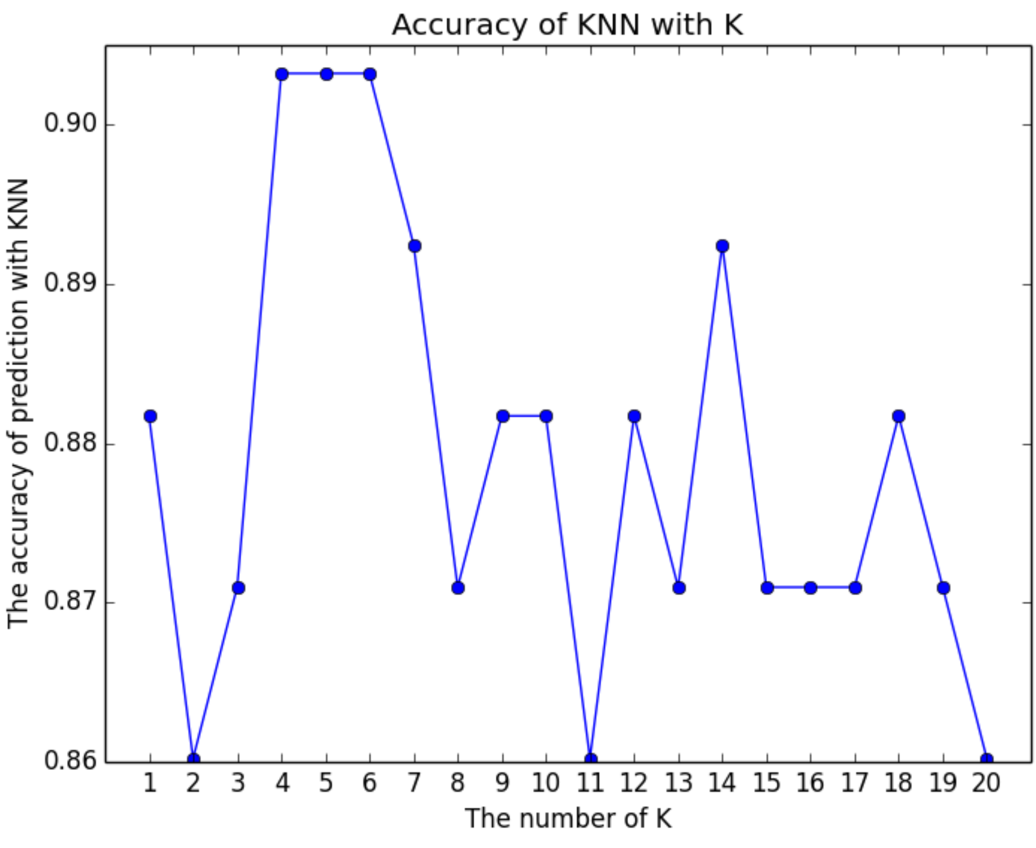






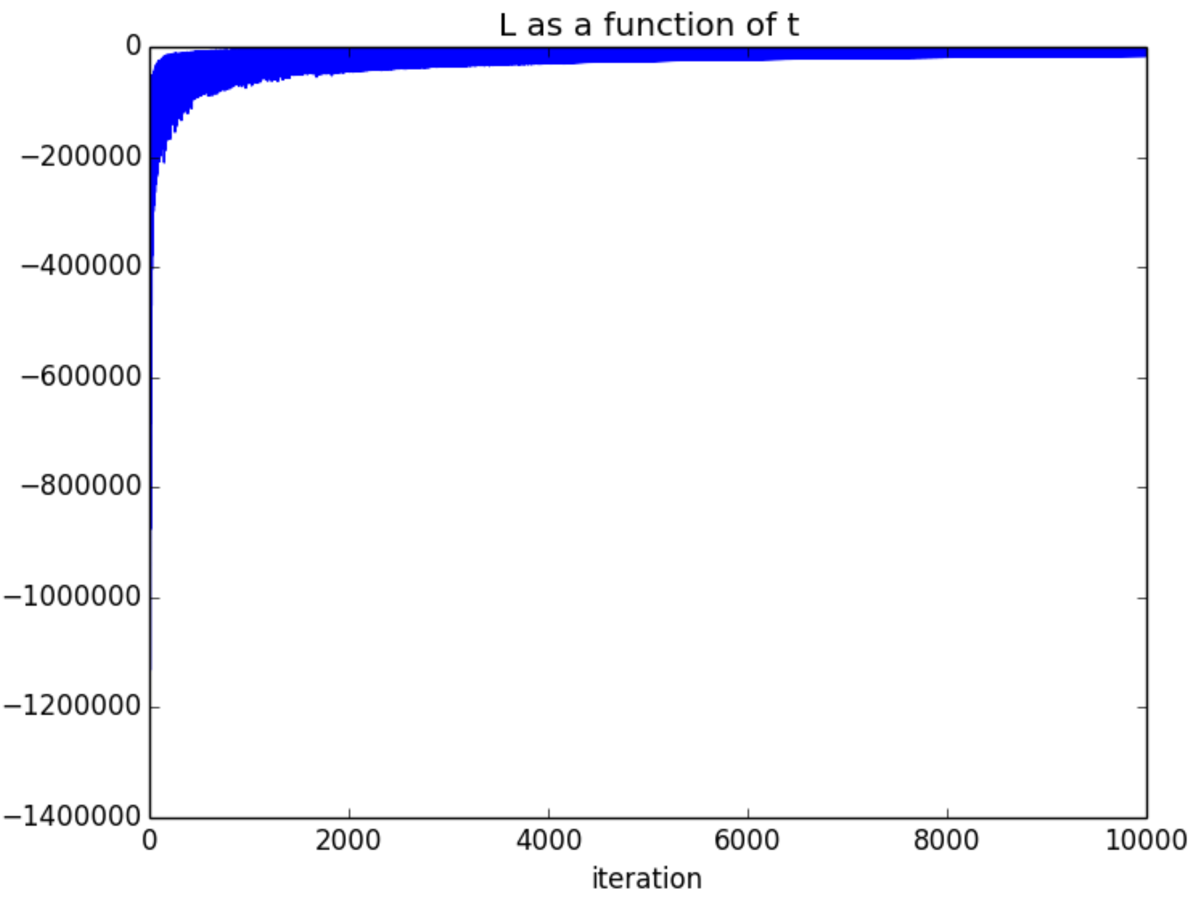
The 16th dimension is the frequency of 'free', and the 52nd dimension is the frequency of char "!". $\theta_1^{(16)} = 0.545$, $\theta_1^{(52)} = 0.833$, $\theta_0^{(16)} = 0.091$, $\theta_0^{(52)} = 0.269$. In calculation, the value of $\theta_{y_i}^{(d)}$ is the mean of the d-th feature given that $y = y_i$. We are using MLE, so if we assume that the training data is sufficient, then the value of $\theta_{y_i}^{(d)}$ is also the probability that word 'free' or char "!" appear given that $y = y_i$ (In Naive Bayes classifier, this two features are independent). For example, the value of $\theta_1^{(52)}$ is high, that means the character '!' are likely to appear in a spam email. On the other hand, the value of $\theta_0^{(16)}$ is quite small, which means the word 'free' are not likely to appear in a normal email.

c. The accuracy of KNN with k from 1 to 20 is as follow.



d.

The pattern do look strange.



e.

The graph of objective function as a function of t is as follow. The prediction accuracy is about 0.91398.