CSCI567 2013 Homework Assignment 3

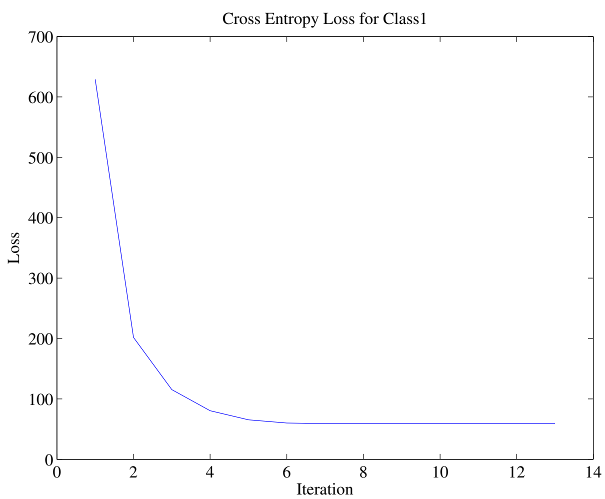
Programming Report

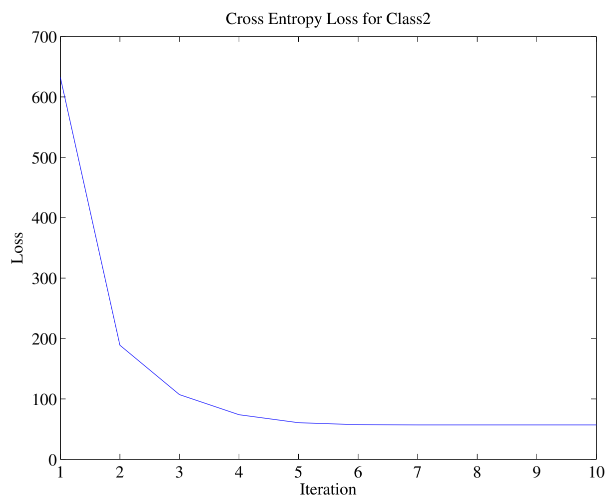
Student Name	Last 4 digits of USC ID
Jia Li	0854
Udit Agrawal	5165
Yin-Ray Rick Huang	6794
Arpit Bansal	0979

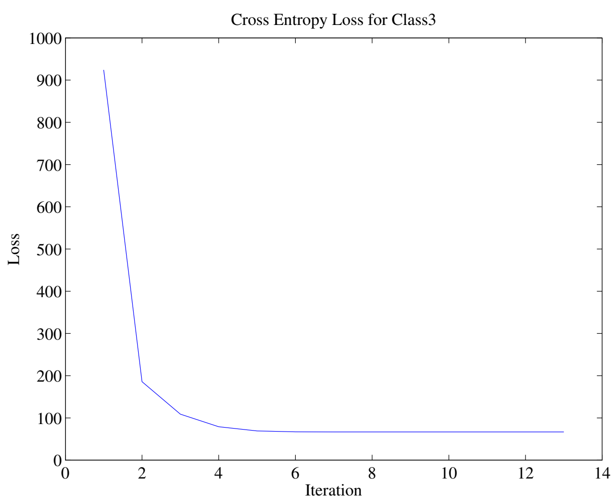
1 Regularized Cross Entropy

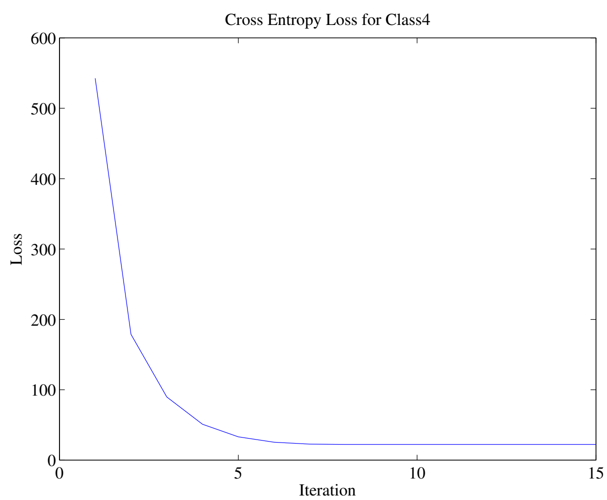
T = 100

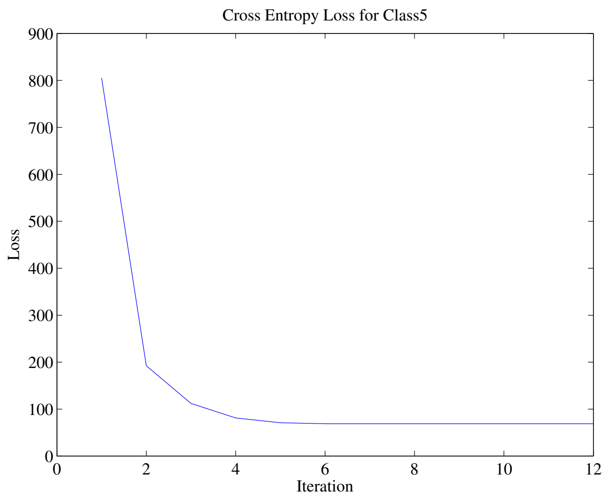
Class ID	1	2	3	4	5	6	7	8	9	10
Best Lambda	90000	70000	120000	30000	120000	30000	10000	160000	70000	10000

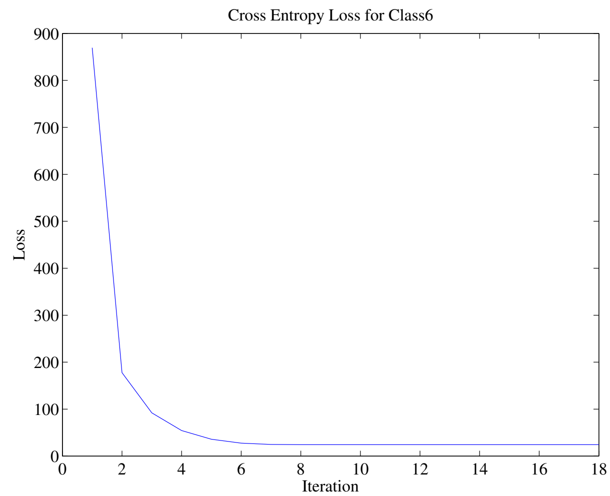


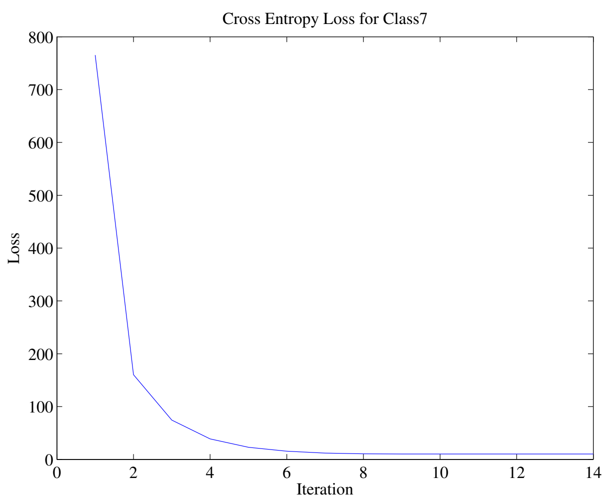


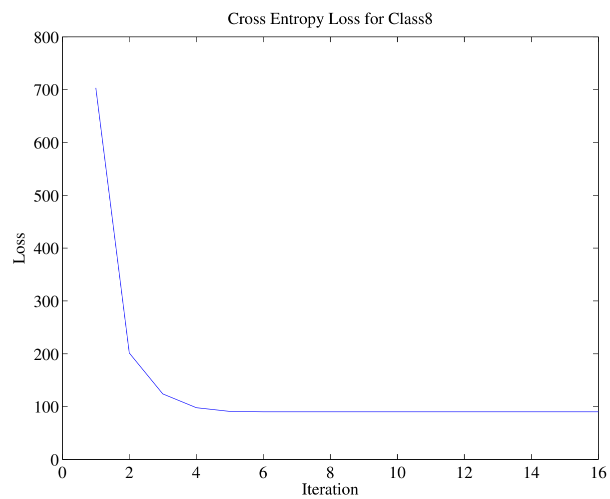


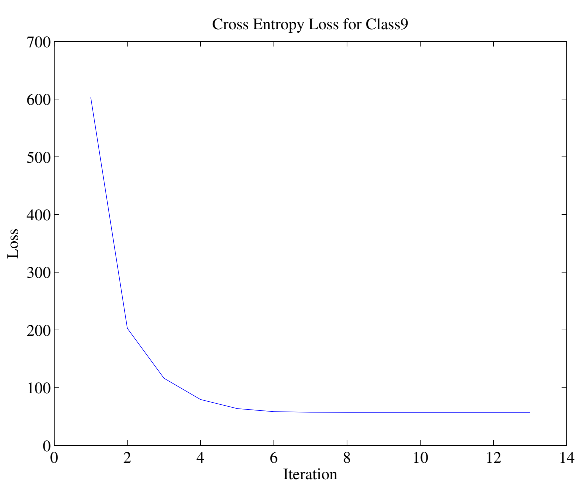


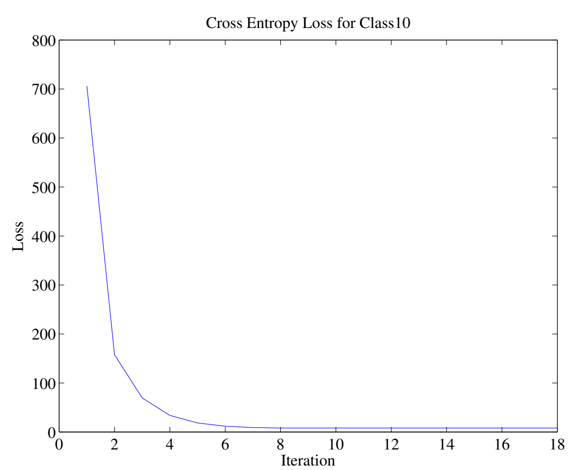

Cross Entropy Loss for Class5
Loss
Iteration


Cross Entropy Loss for Class6
Loss
Iteration


Cross Entropy Loss for Class7
Loss
Iteration


Cross Entropy Loss for Class8
Loss
Iteration


Cross Entropy Loss for Class9
Loss
Iteration


Cross Entropy Loss for Class10
Loss
Iteration

2 Accuracy

2.1 One Vesus Rest

Training Accuracy = 0.9920
Testing Accuracy = 0.9030

2.2 One Vesus One

Training Accuracy = 1
Testing Accuracy = 0.9340

3 Combine One-vs-One

To combine the one-vs-one result, we used something like the PageRank. After we get all the one-vs-one result, we can treat all the 10 classes as node in the graph. This is fully connected graph, each two node on the graph are connected by a edge. The classifier like $i - vs - j$ we trained from the logistic regression is, we model the $p(i)$ as the probability that some energy will flow from node j to node i, and reverse probability of flow (from node j to node i) is $1 - p(i)$. Also for each node, it has a probability 1 to stay on itself. (Don't worry about the one, we will normalize these things). Then for each node, we will have 10 out flow path (one path is to itself), we normalized these 10 probability by their sum.
Now we got a transforming matrix $\boldsymbol{T}$, and each element $T_{i,j}$ stands for the probability that energy will flow from node j to node i.

After building the transforming matrix, then we can start from a initial vector $v = [0.1, 0.1...0.1]^T$ having 10 elements with each element value 0.1. Then for the vector v, we need to multiply T to it on left hand side for multiple times(10 times will already converge). After doing that, we just select the index of $\boldsymbol{v}$which has the largest value as the label.

4 Combine One-vs-Rest

This one is really trivial, as we have ten classifiers, and for each testing sample, we will have ten probability values whether it belongs to that class. We just need to select the class that has the largest probability value.

5 Extra Credit

5.1 Accuracy

Training Accuracy = 0.9830
Testing Accuracy = 0.8385

5.2 Closed Form Solution

$$
\begin{aligned}
\pi_k &= \sum_n \mathbb{1}(y_n = k) \\
\boldsymbol{\mu_k} &= \frac{\sum_n \mathbb{1}(y_n = k)\boldsymbol{x}_n}{\sum_n \mathbb{1}(y_n = k)} \\
\boldsymbol{\Sigma} &= \frac{1}{N}\sum_n \sum_k \mathbb{1}(y_n = k)(\boldsymbol{x}_n - \boldsymbol{\mu_k}) \\
\boldsymbol{\omega_k} &= \boldsymbol{\Sigma^{-1}}\boldsymbol{\mu_k} \\
b_k &= -\frac{1}{2}\boldsymbol{\mu_k^T}\boldsymbol{\Sigma^{-1}}\boldsymbol{\mu_k} + \log \boldsymbol{\pi_k} \\
p(y = k|\boldsymbol{x}, \omega, b) &\propto e^{\omega_k \boldsymbol{x} + b}
\end{aligned}
$$