CS 615 - Deep Learning

Assignment 1 - Artificial Neurons
Spring 2020

Xiangang Lai

1 Theory

1. Given a class probability distribution of $\{0.1, 0.8, 0.1\}$, what is the cross entropy if the correct class is the first class? (3pts)

$$J = -ln(ya) = -ln(0.1) = 2.3026$$

2. If we're using cross-entropy as our objective function, are we attempting to minimize it or maximize it (2pts)?
Minimize, to reduce the randomness.

3. Given the confusion table in Figure 1:

		True Class			
		1	2	3	4
Predicted Class	1	5	2	3	4
	2	8	12	30	4
	3	0	8	45	4
	4	10	0	5	80

Figure 1: Confusion Matrix

- What are the class priors? (3pts)
The prior was calculated based on the number of true classes over the total numbers. get:
0.1045 0.1000 0.3773 0.4182
- What is the overall accuracy of the system? (2pts)
The accuracy equal the sum of diagonal entries over the total number. Gives:
0.6455

2 Gradient Descent

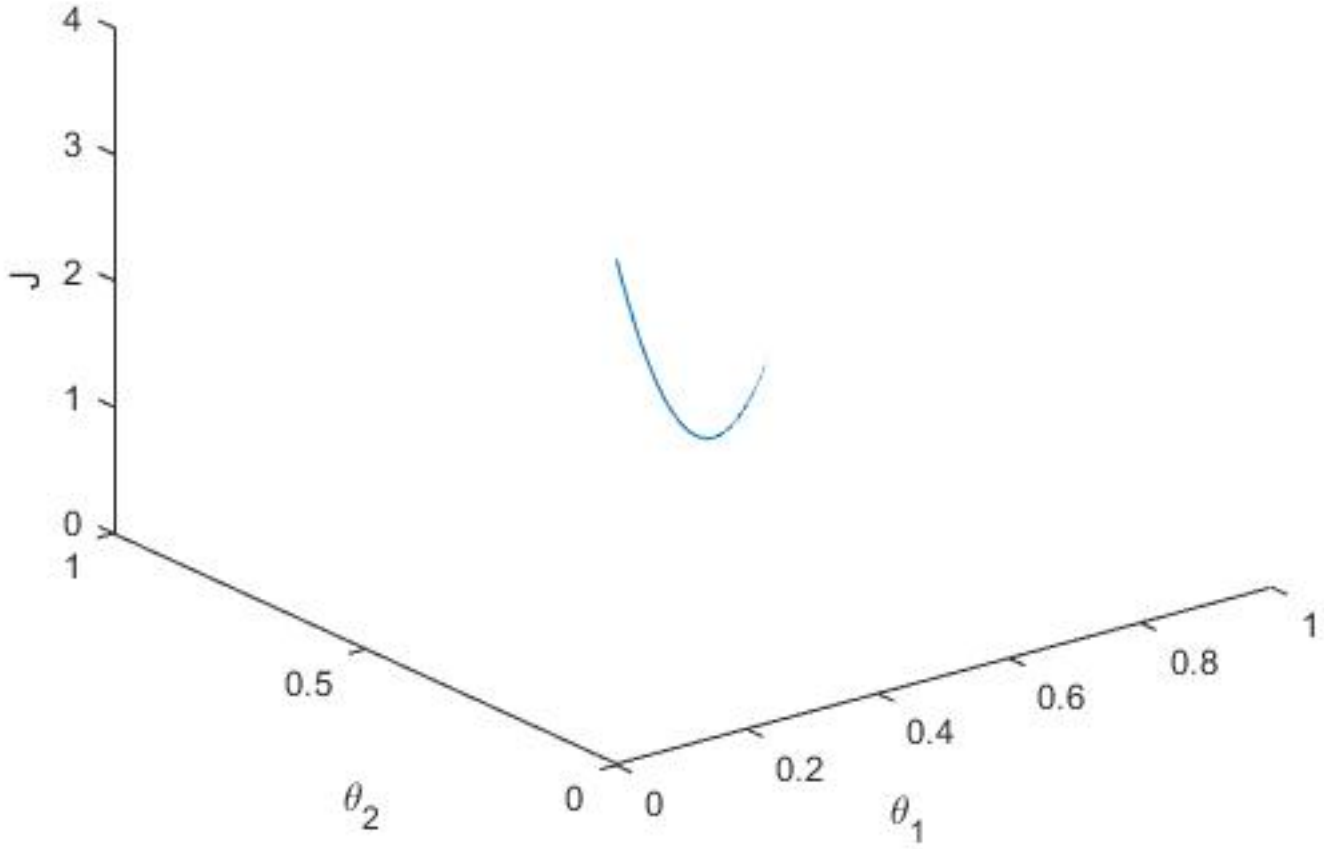

Figure 2: Gradient Descent Progress

3 Gradient Descent Logistic Regression

1. The values of the hyperparameters that you chose.

   - theta was initialized as a rand matrix with dimension 1601x14, with seed rng(1)
   - learning rate: 0.1
   - L2 regularization ratio: 0.01
   - termination criterion: reach 1000 iterations, or change in objective function J or norm of change of theta smaller than 1e-6, which one arrives earlier

2. A graph of the average log likelihood for the training and testing sets as a function of the training iteration number.

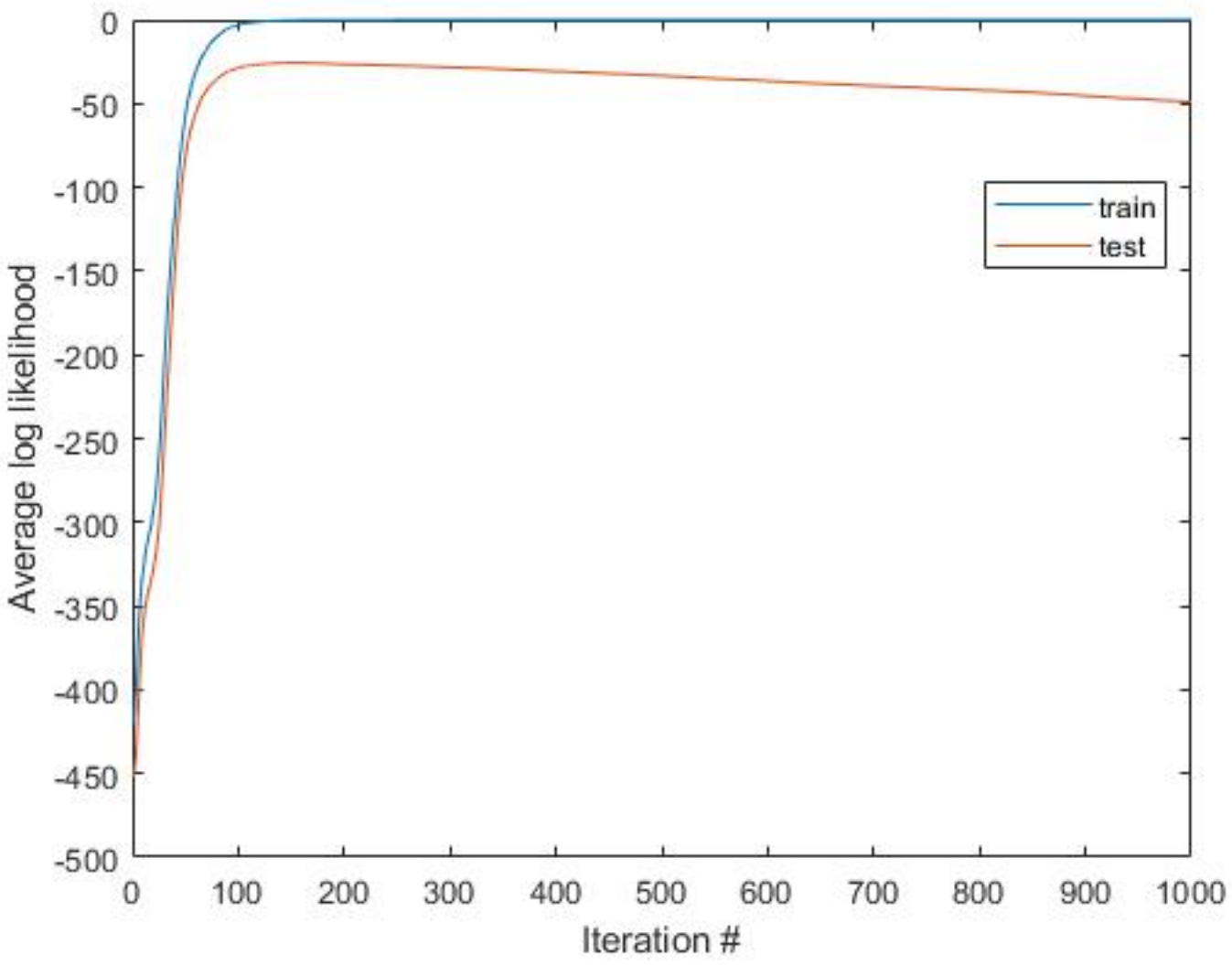

Figure 3: Gradient Descent Logistic Regression Progress

3. The testing accuracy.
   0.9286

4. The confusion matrix for the testing data.

		True Class													
		2	3	4	5	6	7	8	9	10	11	12	13	14	15
Predicted class	2	4	0	0	1	0	0	0	0	0	0	0	0	1	0
	3	0	3	0	0	1	1	0	0	0	0	0	0	1	0
	4	1	0	4	1	0	0	0	0	0	0	0	0	0	1
	5	0	0	1	4	1	0	0	0	0	0	0	0	1	1
	6	0	0	0	0	4	0	0	0	0	0	0	0	2	0
	7	0	0	1	0	0	4	0	0	0	0	0	0	0	0
	8	0	1	0	0	0	0	2	1	1	1	0	0	1	0
	9	1	0	0	1	0	0	0	4	0	0	0	0	1	0
	10	0	0	1	0	0	0	1	0	4	0	0	0	0	0
	11	0	0	1	0	0	0	0	0	0	4	1	0	1	1
	12	0	0	0	0	0	0	0	0	0	0	4	0	1	0
	13	0	0	0	0	0	0	1	0	0	0	0	4	1	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	4	0
	15	0	0	0	0	0	1	0	0	0	0	0	0	0	3

4 Gradient Descent w/ Softmax and Cross-Entropy

1. The values of the hyperparameters that you chose.
   - theta was initialized as a zero matrix with dimension 1601x14
   - Learning rate: 0.08
   - L2 regularization ratio: 0.85
   - Termination criterion: reach 1000 iterations, or change in objective function J or maximum column norm of change of theta smaller than 1e-6, which one arrives earlier
2. A graph of the average cross-entropy loss the training and testing sets as a function of the training iteration number.

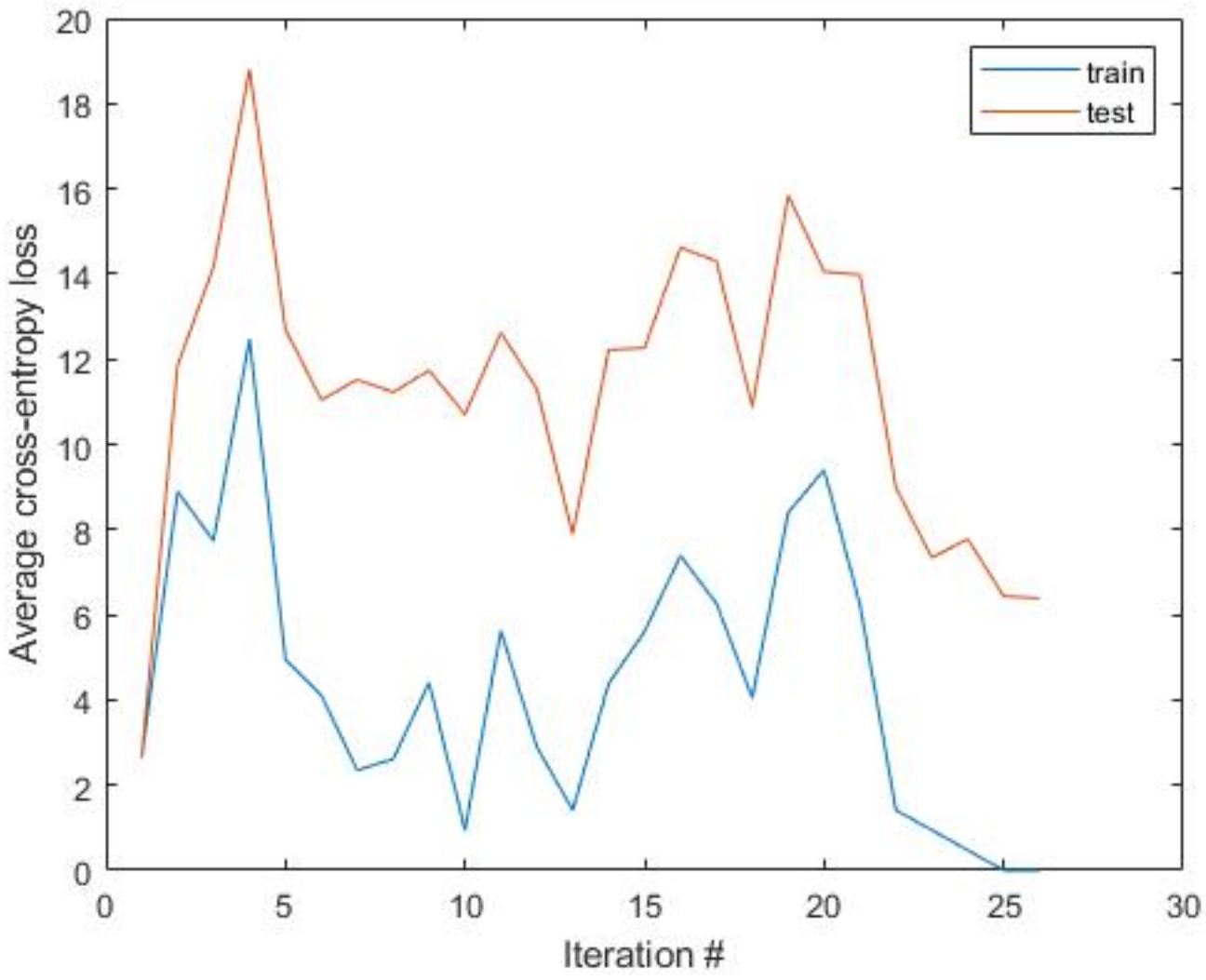

Figure 4: Gradient Descent Logistic Regression Progress

3. The testing accuracy.

   0.9286

4. The confusion matrix for the testing data.

		True Class													
		2	3	4	5	6	7	8	9	10	11	12	13	14	15
Predicted class	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	4	0	0	0	0	0	0	0	0	0	0	0
	5	1	0	0	4	0	0	0	0	0	0	0	0	0	1
	6	0	0	0	0	4	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	4	1	0	1	0	0	1	0	1
	8	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	4	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	3	0	0	0	0	0
	11	0	1	0	0	0	0	0	0	0	4	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	4	0	0	0
	13	0	0	0	0	0	0	1	0	0	0	0	3	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	4	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	2