

ENGR 421 / DASC 521: Introduction to Machine Learning

Homework 02: Discrimination by Regression

Deadline: March 27, 2021, 11:59 PM

In this homework, you will implement a discrimination by regression algorithm for multiclass classification in Python. Here are the steps you need to follow:

1. Read Section 10.7.3 (4th edition) or Section 10.8 (3rd edition) from the textbook.
2. You are given a multivariate classification data set, which contains 1000 clothing images of size 28 pixels \times 28 pixels (i.e., 784 pixels). These images are from five distinct classes, namely, T-shirt, trouser, dress, sneaker, and bag. The figure below shows five sample figures from each class. You are given two data files:
 - a. `hw02_images.csv`: letter images,
 - b. `hw02_labels.csv`: corresponding image labels (1: T-shirt, 2: trouser, 3: dress, 4: sneaker, and 5: bag).

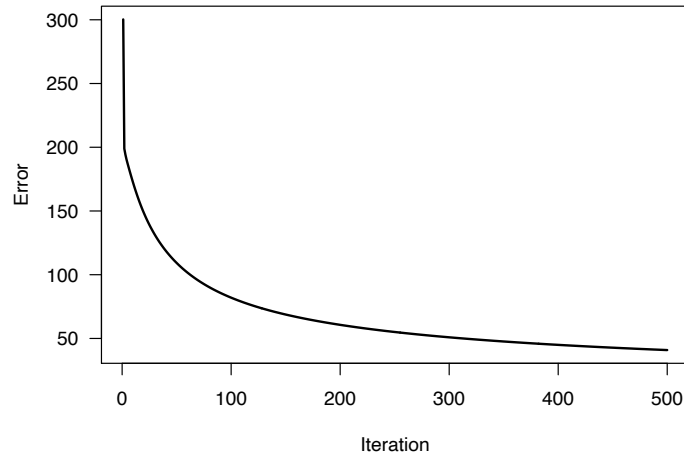


3. Divide the data set into two parts by assigning the first 500 images to the training set and the remaining 500 images to the test set.
4. Learn a discrimination by regression algorithm using the sigmoid function for this multiclass classification problem. You can use the following learning parameters.

```
eta = 0.0001
epsilon = 1e-3
max_iteration = 500
```

You should initialize the weight parameters to the values given in `initial_W.csv` and `initial_w0.csv` files before running your algorithm.

5. Draw the objective function values throughout the iterations. Your figure should be similar to the following figure.



6. Calculate the confusion matrix for the data points in your training set using the discrimination rule you will develop using the estimated parameters. Your confusion matrix should be similar to the following matrix.

		y_train				
y_predicted		1	2	3	4	5
1	101	2	11	0	1	
2	0	79	1	0	0	
3	6	5	100	0	1	
4	0	0	0	99	2	
5	0	0	0	1	91	

7. Calculate the confusion matrix for the data points in your test set using the parametric discrimination rule you will develop using the estimated parameters. Your confusion matrix should be similar to the following matrix.

		y_test				
y_predicted		1	2	3	4	5
1	82	1	4	0	2	
2	0	87	0	0	0	
3	13	3	98	0	3	
4	0	0	0	103	1	
5	4	0	3	0	96	

What to submit: You need to submit your source code in a single file (.py file) and a short report explaining your approach (.doc, .docx, or .pdf file). You will put these two files in a single

zip file named as ***STUDENTID.zip***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the zip file you created to Blackboard. Please follow the exact style mentioned and do not send a zip file named as ***STUDENTID.zip***. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.