

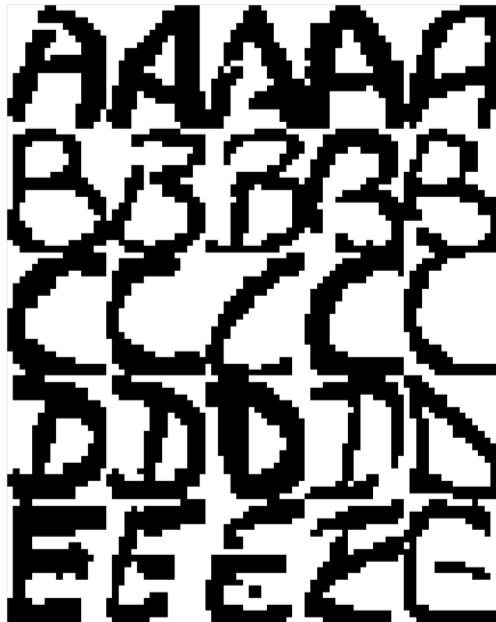
ENGR 421 / DASC 521: Introduction to Machine Learning

Homework 03: Discrimination by Regression

Deadline: March 28, 2022, 11:59 PM

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

1. Your discrimination by regression algorithm will be developed using the following modifications to the linear discrimination algorithm with the softmax function we discussed in the lectures.
 - a. Instead of the softmax function, you are going to use K sigmoid functions to generate \hat{y}_{ic} values. Please note that, in such a case, the summation of \hat{y}_{ic} values is not guaranteed to be 1. However, you are going to pick the largest value to predict the class label.
 - b. Instead of the negative log-likelihood (i.e., $-\sum_{i=1}^N \sum_{c=1}^K y_{ic} \log \hat{y}_{ic}$), you are going to use the sum squared errors as the error function to minimize (i.e., $0.5 \sum_{i=1}^N \sum_{c=1}^K (y_{ic} - \hat{y}_{ic})^2$). Please note that you need to find the correct update equations for this modified model.
2. You are given a multivariate classification data set, which contains 195 handwritten letters of size 20 pixels \times 16 pixels (i.e., 320 pixels). These images are from five distinct classes, namely, A, B, C, D, and E, where we have 39 data points from each class. The figure below shows five sample figures from each class. You are given two data files:
 - a. `hw03_data_set_images.csv`: letter images,
 - b. `hw03_data_set_labels.csv`: corresponding class labels.



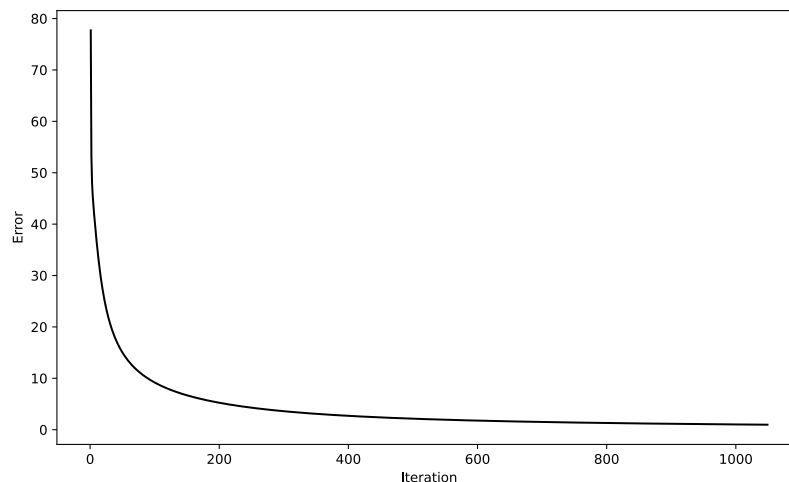
3. Divide the data set into two parts by assigning the first 25 images from each class to the training set and the remaining 14 images to the test set. (10 points)
4. Learn a discrimination by regression algorithm for this multiclass classification problem using the following learning parameters.

```
eta = 0.001  
epsilon = 0.001
```

Your parameter estimations should be like the following figures. (40 points)

```
print(W)  
print(w0)  
[[-0.00728158 -0.0471197 -0.00647147  0.11842513 -0.02009192]  
 [-0.07200798  0.01709564 -0.12547436  0.19233187  0.03607319]  
 [-0.07017481  0.05245197 -0.12845519  0.14109499  0.03367804]  
 ...  
 [ 0.02956927  0.13179479  0.04025555 -0.13815584 -0.12179576]  
 [ 0.05512022 -0.15812092  0.04773618 -0.18309446 -0.02013914]  
 [ 0.07311371 -0.24000196  0.08470467 -0.14902205 -0.01617009]]  
[[-0.05081352 -0.16541552 -0.00299109 -0.03717156 -0.15866353]]
```

5. Draw the objective function values throughout the iteration. Your figure should be like the following figure. (10 points)



6. Calculate the confusion matrix for the data points in your training set using the discrimination by regression algorithm you will develop using the estimated parameters. Your confusion matrix should be like the following matrix. (20 points)

```
print(confusion_matrix)
y_truth  1   2   3   4   5
y_pred
1         25   0   0   0   0
2         0  25   0   0   0
3         0   0  25   0   0
4         0   0   0  25   0
5         0   0   0   0  25
```

7. Calculate the confusion matrix for the data points in your test set using the discrimination by regression algorithm you will develop using the estimated parameters. Your confusion matrix should be like the following matrix. (20 points)

```
print(confusion_matrix)
y_truth  1   2   3   4   5
y_pred
1         13   1   0   0   0
2         1  11   0   0   2
3         0   0  14   0   0
4         0   0   0  14   0
5         0   2   0   0  12
```

What to submit: You need to submit your source code in a single file (.py file) named as ***STUDENTID.py***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the file you created to Blackboard. Please follow the exact style mentioned and do not send a file named as ***STUDENTID.py***. 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.