MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY, JAIPUR

Department of Computer Science and Engineering

M.Tech Programming Lab 2019

Assignment-1

Note:
(1) All the problems are to be implemented in Python. The exercises are to be done individually.
(2) The submission date is the last lab day for the exercise. The number of labs for each exercise is mentioned alongside.

Following is a list of exercises to be done as part of Machine Learning Lab course.

1. (15 points, Labs - 2) Handwritten Digits Data: You should download the two data files with handwritten digits data: training data (ZipDigits.train) and test data (ZipDigits.test). Each row is a data example. The first entry is the digit, and the next 256 are grayscale values between −1 and 1. The 256 pixels correspond to a 16 × 16 image. For this problem, we will only use the 1 and 4 digits, so remove the other digits from your training and test examples. Please submit your Python code implementing the logistic regression for classification using gradient descent.

    1. (5 points) Familiarize yourself with the data by giving a plot of two of the digit images.
    2. (5 points) Develop two features to measure properties of the image that would be useful in distinguishing between 1 and 4. You may use symmetry and average intensity (as discussed in class).
    3. (5 points) As in the text, give a 2-D scatter plot of your features: for each data example, plot the two features with a red redx if it is a 4 and a blue blueo if it is a 1.
    4. (35 points) Classifying Handwritten Digits: 1 vs. 4. Implement logistic regression for classification using gradient descent to find the best separator you can using the training data only (use your 2 features from the above question as the inputs). The output is +1 if the example is a 1 and −1 for a 4.
    5. (5 points) Give separate plots of the training and test data, together with the separators.
    6. (10 points) Compute E_{in} on your training data and E_{test}, the test error on the test data after 1000 iterations.
    7. (10 points) Now repeat the above using a 3rd order polynomial transform.
    8. (10 points) As your final deliverable to a customer, would you use the linear model with or without the 3rd order polynomial transform? Explain.