CSE 404: Introduction to Machine Learning (Fall 2020)

Homework #7

Due 11/16/2020 by 11.59pm

Note: (1) Please upload a soft copy of your homework on D2L. Please type your answers. (2) Please upload the your Python code separately, do NOT upload a zip folder.

1. (40 points) Handwritten Digits Data: You should download the data files with handwritten digits data including only 1 and 5: training data (train_data.npy), training labels (train_labels.npy), test data (test_data.npy), and test labels (test_labels.npy). You can use np.load() to load the npy files. Each row of train_data and test_data represents one data point. train_data should be a 1561×256 matrix and test_data should be a 424×256 matrix. Each data point has 256 gray scale values between -1 and 1. The 256 pixels correspond to a 16×16 image. train_labels and test_labels are 1561 and 424 dimensional arrays, respectively and they have label 1 for digit 1 and label -1 for the digit 5.

   (a) (10 points) Plot of two of the digit images, one for digit 1 and one for digit 5.

   (b) (20 points) Extract the two features discussed in the class (symmetry and average intensity) to distinguish 1 and 5.

   (c) (10 points) Provide 2d scatter plots of your features for training and test data (Now your data matrix will be N×2). For each data example, plot the two features with a red $\times$ if it is a 5 and a blue $\circ$ if it is a 1.

2. (60 points) Classifying Handwritten Digits: 1 vs. 5. Implement logistic regression for classification using gradient descent to find the best linear 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 5.

   (a) (15 points) Give separate plots of the training and test data, together with the separators. (Similar what you did in PLA homework. After you learn the model vector w, you can plot a line. You may want to concatenate 1 to your data for the intercept term.)

   (b) (15 points) Compute train E_{in} and test E_{test} errors. Use only the training data to compute training error and use only the test data to compute the test error.

   (c) (15 points) Logistic regression can also have regularization: $\min_{\mathbf{w}} E(\mathbf{w}) + \lambda \|\mathbf{w}\|_2^2$, where $E(\mathbf{w})$ is the logistic loss. Change your gradient descent algorithm accordingly and repeat (b). Report the best λ using cross-validation.

   (d) (15 points) Now repeat (b) using a 3rd order polynomial transform.

   (e) (15 points) As your final deliverable to a customer, would you use the linear model with or without the 3rd order polynomial transform? Explain.