Assignment 8

Lorik Mucolli, Aavash Shrestha

Linear Classifier

The classifier we have used for this assignment is a complete linear classifier. That is, the whole process of feature extraction and classification we have applied on the digits data could be performed by a single matrix. We, however, have done it step by step. The main classifier is implemented in *classifier.py* .

Feature Extraction

Principal component analysis(PCA) is used as the pre-processing stage to reduce the dimension of the data. This stage is very important for the classification process as it determines the different models for each classifier. The number of principal components(eigen vectors) of the co-variance matrix of the centered data separates each classifier from one another. The features have the bias appended. PCA is implemented in module *pca.py*.

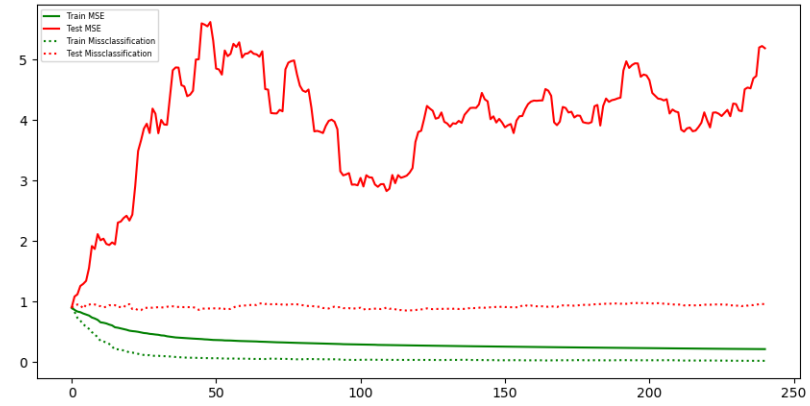
Decision Function Approximation

Linear Regresssion is used to approximate the decision function that takes in the features(obtained from PCA with bias appended) and gives the digits class vectors(binary vectors for class labels). The algorithm given in the lecture notes has been followed. Only training data(first 100 instances of each digit) were used to get the optimal weight matrix. Linear regression is implemented in module *l_reg.py*.

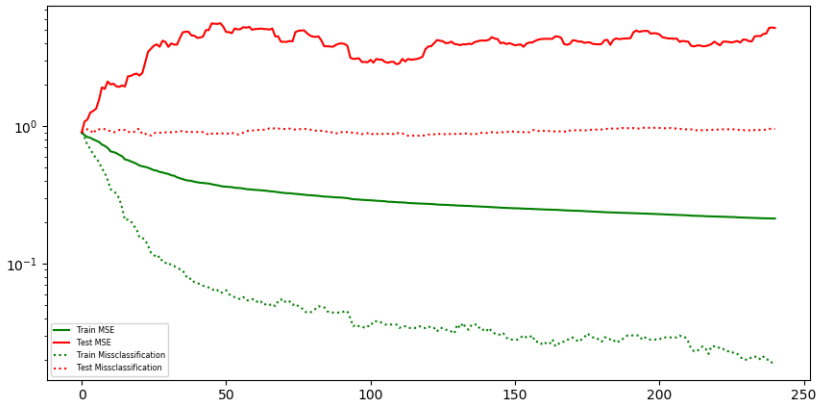
Classification

The optimal weight matrix obtained from linear classification is used to train both the test and the train data(features extracted by PCA first). The following results were obtained for *pcs = 0,1,2,....240*

Results



`Linear scale`



`Logarithmic scale`

The calculation of the mean squared errors (MSE) and the the miss-classification (MISS) rates are in *classifier.py*. The results are very interesting. The graph we obtained is what is expected for the test data. Both the MSE and the MISS for the train data decrease sharply at first as K increases and then decrease slowly for high K as we fit the data more. The miss-classification rate is also significantly low.
What's interesting is the results for the test data. Initially, the MSE seems to rise up as K increases and the it begins to fall again. Then as we begin to

overfit the data, by using more principal components, the MSE goes up again(as expected). The test MISS, however, is stable throughout the graph, but it is much higher than the train MISS.
We believe this inconsistent result for the MSE is because we are using PCA for feature extraction. PCA is linear, and so, it fails to capture the non-linear correlations in the data, specially for the digits sample we have. Our linear classifier fails to capture the non-linear manifolds that the data and correlations actually reside in. In conclusion(empirical), a linear classifier would not be best for the digits data.