Machine Learning Final Project

103062512 徐丞裕

1. Training Error:

   16% (8/50) over the labeled training instances.

2. Testing Error:

   10% (10/50) over the given testing data.

3. Algorithm Steps:

   A. Data Preprocessing:

      Perform Z-normalization on the training instances.

   B. Label Propagation:

      Use Laplacian Support Vector Machine (LapSVM) on the training instances to learn the labels of the unlabeled data.

   C. Feature Selection:

      Use Linear Discriminant Analysis (LDA) on the training data and the labels learned from B. to find the projection vectors.

      We then filter out insignificant features using a threshold and reduce the dimensions of the training data.

   D. Train the Classifier:

      Train the final classifier using the data we collected from the

previous steps (i.e. labeled learned, instances with dimensions reduced).

4. Implementation:

   A. Use build-in function `zscore` to perform z-normalization.

   B. We used the following objective to train the LapSVM:

$$\min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} \left(1 - y_i f(x_i)\right) + \lambda \|f\|_K^2 + \frac{\mu}{n^2} f^T L f$$

   And solve the dual problem extracted from:

   Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples (Mikhail Belkin, 2006).

   C. Implement LDA algorithm from the lecture notes and filter out the features whose absolute values are less than 0.4.

   D. Use the LapSVM model and set the μ to zero (result in a tradition SVM) to train a classifier over the training instances and the labels we obtain at step B. Also, a 8-fold cross validation is used to choose the best model under the same hyper-parameters.