a. Two semi supervised learning for LDA

1) Method 1

Inspired by reference [1], the first self-training-based approach using semi-supervised linear discriminant analysis is implemented in this answer sheet. The algorithmic procedure is listed as below

Input: Labeled set, sets containing labels, unlabeled set,

Output: A projection matrix W (eigenvector) mapping the data onto lower-dimension space

Method:

1. Train the labeled data via LDA
2. Apply extracted projection matrix on unlabeled data to obtain predicted labels
3. Add new unlabeled data to current data sets to update the projection matrix in LDA
4. Loop step 3 and keep updating the LDA within each iteration until total iterations>100 && variance of the predicted labels between two adjacent iteration <10^{-5}.
5. Advance the classifier to the latest LDA after convergence

2) Method 2

The second method implements also semi-supervised LDA approach but synthesize moment constrains simultaneously [2]. It constrains the possible configurations that the class means can take on by linking label dependent parameters with label independent parameters, mapping the combined labeled data sets and unlabeled sets onto a new set with same dimension as the labeled data, which contains both the information from labeled and unlabeled data but has lower dimension. The algorithm can be elaborated as follows,

Input: Labeled set, sets containing labels, unlabeled set,

Output: A projection matrix W (eigenvector) mapping the data onto lower-dimension space

Method:

1. Compute the means of the two classes as μ_1 and μ_2 respectively
2. Compute the means of the combined data sets and labeled data sets as μ_T and μ_l
3. Compute the covariance matrix of labeled data sets and total data sets as Θ_l and Θ_T
4. Transform the current space with labeled and unlabeled data to the sub-space through $x' = \Theta_T^{\frac{1}{2}} \Theta_l^{-\frac{1}{2}} \left(x - \mu_l \right) + \mu_T$
5. Train the new sub-space which has the same dimension as the labeled data and obtain the updated LDA

b. Learning curve as a function of unlabeled data with different methods

In this experiment, the labeled data is randomly picked from the whole space and the unlabeled data is also randomly selected from the space excluding the labeled data. Three different methods, involving supervised training LDA, and two semi-supervised training LDA are implemented on the magic gamma telescope data set. The training sets contain only the first 25 selected labels and the test sets are composed of 2000 randomly selected data from the original data set. To eliminate the numerical error due to randomness, the experiment is repeated 50 times (take the mean)to achieve smoothness for each curve. The unlabeled data added for each iteration is 0, 10, 20, 40, 80, 160, 320 and 640 respectively. The error rates as a function of unlabeled data can be plotted as follows,

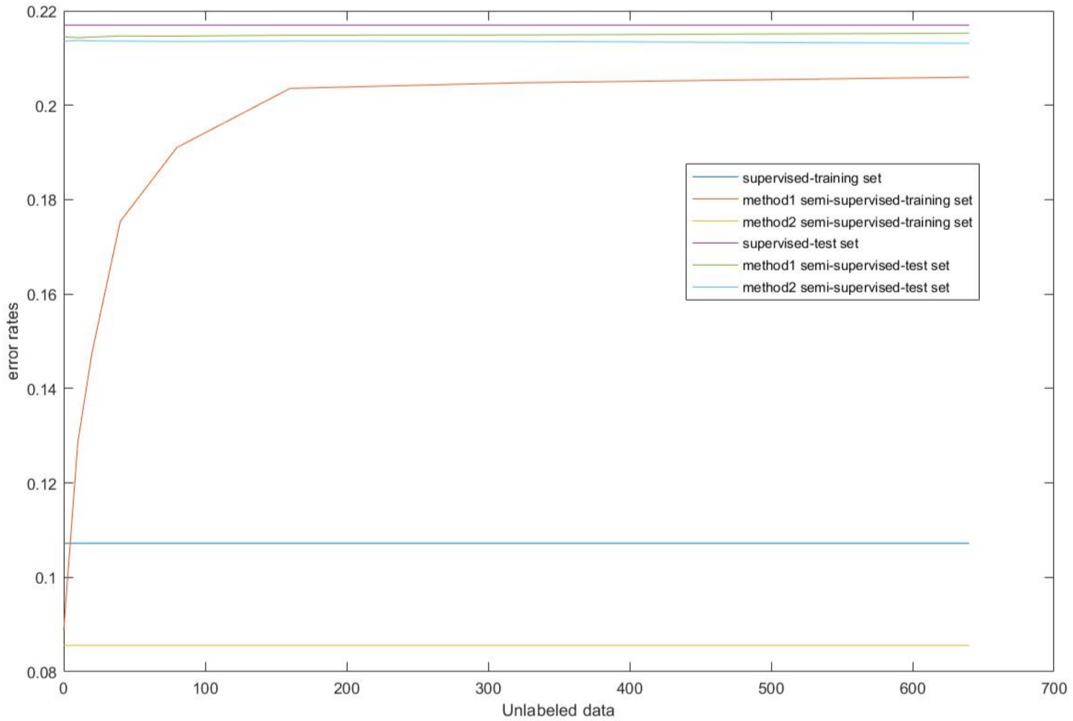

Figure 1 Errors vs unlabeled data

It can be observed from the graph that method 1 performs the worst on the training sets, while the supervised benchmarks behave better and method 2 remains in a relatively low level of error rates. Only negligible oscillations are observed in the error rates, leading to the observation that generally simply changing the number of unlabeled points will not significantly affect the error rates for the LDA classifier when the experiment is repeated enough times. For the test sets, method 2 and method 1 semi-supervised LDA are slightly more accurate than the supervised LDA, but the error rates from all these methods are higher than those in training sets.

c. log-likelihood as a function of unlabeled data with different methods

The log-likelihood vs the number of unlabeled data is plotted in the following figure with different methods implemented.

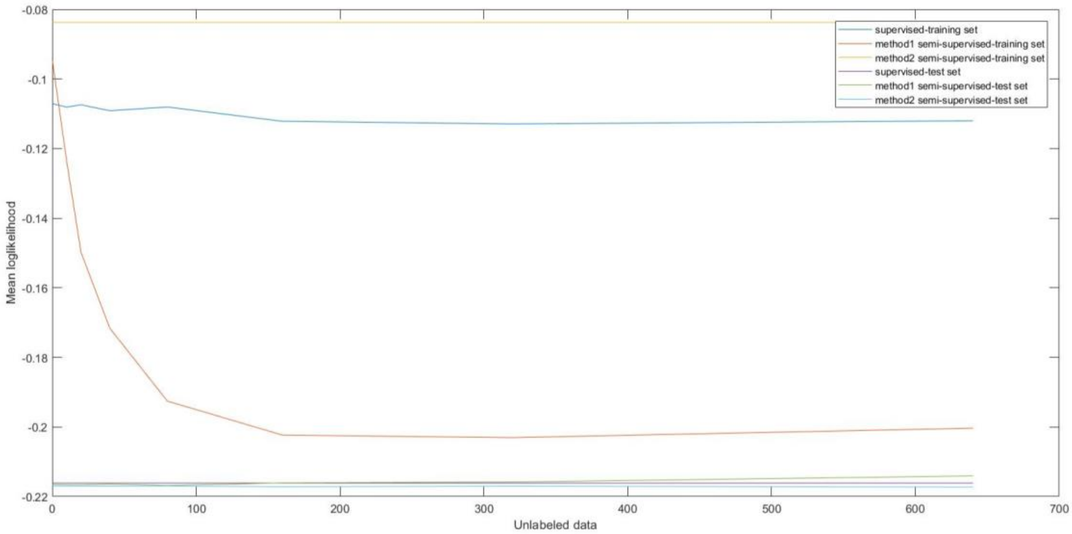

Figure 2 Mean log-likelihood vs number of unlabeled data

d. Construct two artificial data sets.

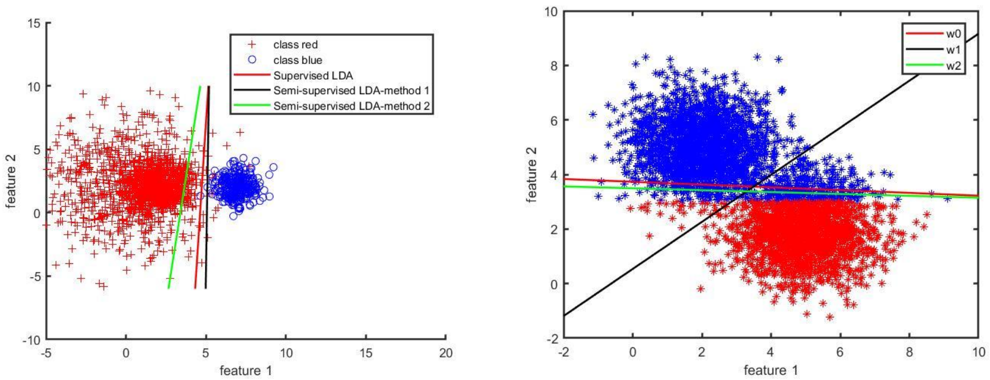

Figure 3 Data set 1 and 2 construction

In the first data set, the error rate for supervised LDA, Method1-Semi-supervised LDA and Method-2 Semi-supervised LDA is 0.0167, 0.0117 and 0.0467, respectively. Obviously, Method-1-Semi-supervised LDA performs the best, while supervised LDA ranks 2^{nd} and Method-2 Semi-supervised LDA ranks 3^{rd}. For simplicity, the data set in this case is composed of 1700 red points and 300 blue points with Gaussian distribution on a 2-D space. The sampling rate for the unlabeled data is 0.2. The covariance matrix for red class has more scattering characteristic and generally these two classes are linearly separated as the distance between their means remains at a high level. The purpose of these settings are to maximize the between-class variance to maximize the influence of Method-1-Semi-supervised LDA classifier. This is because of the fact that method 1 is an error-rate based self-learner, which is ideal for linearly separated cases.

In the second data set, the error rate for supervised LDA, Method1-Semi-supervised LDA and Method-2 Semi-supervised LDA is 0.04, 0.1933, and 0.0267, respectively. The data sets contain 2000 data points for each class with Gaussian distribution. The number of the labeled data is set as 75, to approximate the relations between labeled data and unlabeled in a). The purpose of constructing this data set is to target at the downsides of method-1, which depends on error rate to update the classifier until the convergence. If the data sets are symmetric, method cannot perform well as the error rate will not vary within each iteration. However, the method-2 transforms information of covariance matrix and means of the data including unlabeled ones into the updated projection matrix, which is more suitable to mimic the global behavior in the symmetric case.

Reference

[1] Gan, H., Sang, N., & Huang, R. (2014). Self-training-based face recognition using semi-supervised linear discriminant analysis and affinity propagation. *JOSA A*, *31*(1), 1-6.

[2] Loog, M. (2011, September). Semi-supervised linear discriminant analysis using moment constraints. In *IAPR International Workshop on Partially Supervised Learning* (pp. 32-41). Springer, Berlin, Heidelberg.