HW 3 _LE MINH DONG

Due Oct. 11

## 1. Data information

- Data_train: 32036 x 2049
- Data_test: 8009 x 2049
- $2049^{th}$ column of Data_train and Data_test are the class numbers of 1 to 10
- Data_train and Data_test contains the feature vectors of 2048 dimension (Nx2048)

## 2. Principal Component Analysis (PCA)

- Estimate the covariance matrix from autocorrelation matrices and mean values from training data.
- Calculate and choose m (100, 500, 1000) largest eigenvalues and m respective eigenvectors.
- Use m eigenvectors to form the transformation matrix.
- Use transformation matrix to project original training data and testing data in new space which contain the feature vectors of m dimension (Nxm)
- Put the parameters of updated training and test data in Euclidian and mahalanobis distance classifiers ( similar to homework 1)
- The result of PCA with given test data is shown in table 1

*Table 1. Results of PCA*

| Projection matrix | Principal Component Analysis (PCA) | | | |
|---|---|---|---|---|
| | Euclidian Classifier | | Mahalanobis Classifier | |
| | Error | Processing Time | Error | Processing Time |
| Nx100 * | 0.864 | 2 minutes | 0.706 | 5 minutes |
| NX500 | 0.854 | 5 minutes | 0.690 | 20 minutes |
| NX1000 | 0.85 | 20 minutes | 0.56 | 65 minutes |

*PCA numbers.

- PCA is an algorithm of dimensionality reduction. It generates set of features from original features in more compact way. Therefore, PCA algorithm reduces the time of processing data when the size of projection matrix are decreased. On the other hand, they also make the loss of information, so the value of error are higher than the case of having larger projection matrix.

## 3. Linear Discriminant Analysis (LDA)

- It is an algorithm of dimensionality reduction which is carried out base on supervised mode.
- Calculate mean values, covariance and priori probabilities of 10 classes from training data.
- Calculate within scatter matrix of 10 classes ($S_w$), calculate dimensional between scatter matrix ($S_b$).
- Choose m(100,500,1000) dominant eigenvectors of the matrix product $S_w^{-1}S_b$
- Use m eigenvectors to form the transformation matrix.
- Use transformation matrix to project original training data and testing data in new space which contain the feature vectors of m dimension (Nxm)
- Put the parameters of updated training and test data in Euclidian and mahalanobis distance classifiers ( similar to homework 1)
- The result of LDA with given test data is shown in table 2

*Table 2. Results of LDA*

| Projection matrix | Linear Discriminant Analysis (LDA) | | | |
|---|---|---|---|---|
| | Euclidian Classifier | | Mahalanobis Classifier | |
| | Error | Processing Time | Error | Processing Time |
| Nx100** | 0.1083 | 3 minutes | 0.069 | 30 minutes |
| NX500 | 0.1083 | 7 minutes | 0.067 | 3 hours |
| NX1000 | 0.1083 | 20 minutes | 0.065 | 7 hours |

** Dominant eigenvector of LDA

- LDA is also an algorithm of dimensionality, however it's result in the accuracy of classification are better than PCA's performance because LDA is carried out in supervised mode. It achieves set of features from original features by imposing the mean values of classes are as far apart as possible and the variances values are as small as possible. This is also the reason make LDA take longer time to process given data set than PCA.