

Term Project for ISA5305

Due by 15:40 Monday, June 12, 2017

This term project is designed for you to further understand PCA, LDA, and Data Clustering as introduced in class by real implementation on some data sets. Moreover, the implementation by Independent Component Analysis (ICA) is also *encouraged* but is *not required*.

- (1) For the data set "8OX" introduced in class, there are $n = 45$ patterns from $k = 3$ categories, each pattern consists of $d = 8$ features. Each pattern can be denoted by $\mathbf{x}_i^{(k)}$, $1 \leq i \leq 15$, $1 \leq k \leq 3$, where $\mathbf{x}_i^{(k)} \in R^d$.
 - (a) Compute the pooled $d \times d$ covariance matrix $C = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i^{(k)} - \mathbf{u})(\mathbf{x}_i^{(k)} - \mathbf{u})^t$, where $\mathbf{u} = \frac{1}{n} \sum_{k=1}^3 \sum_{i=1}^{15} \mathbf{x}_i^{(k)}$ is the mean vector.
 - (b) Report the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ of C .
 - (c) Report the percentage of $\gamma_j = \frac{\sum_{i=1}^j \lambda_i}{\sum_{i=1}^d \lambda_i}$, $\forall 1 \leq j \leq d$.
 - (d) Plot n patterns using the first *two* principal components.
 - (e) Plot n patterns using the first *three* principal components.
 - (f) Plot n patterns using the most *two* discriminative features based on linear discriminant analysis (LDA).
 - (g) Plot n patterns using the most *three* discriminative features based on linear discriminant analysis (LDA).
 - (h) Show the dendrogram (by complete linkage) of the original "8OX" data using the $d = 8$ features.
 - (i) Show the results of K-means implementation on the original "8OX" data using the $d = 8$ features.
- (2) For the data set "iris" introduced in class, there are $n = 150$ patterns from $k = 3$ categories, each pattern consists of $d = 4$ features. Repeat the same processes as required in problem (1).
- (3) Find or collect a data set "mydata.txt" with n patterns, d features, and k categories with $5 \leq d \leq 12$, $30 \leq n \leq 50$, $2 \leq k \leq 4$. Repeat the processes as required in problem (1).