Intelligent Data Analysis

Martin Russell
School of Computer Science
Thursday, 06 February 2020

Exercise sheet – week 4 – Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA)

1. Complete the application of PCA to the investigation of the 'nox' parameter in the Boston data. Check that the projection onto the first two principal components is the same as on the slides
2. Apply LDA to investigate the 'nox' parameter in the Boston data. Project the data onto the first two LDA eigenvectors.
3. Let $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ be the set of two-dimensional vectors defined by:

$\mu_1 = 2, \mu_2 = -1$

$$x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}^{1}, x_2 = \begin{bmatrix} 5 \\ -1 \end{bmatrix}^{-3}, x_3 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}^{8}, x_4 = \begin{bmatrix} 4 \\ -3 \end{bmatrix}^{-4},$$
$$x_5 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}^{-1}, x_6 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}^{0}, x_7 = \begin{bmatrix} 7 \\ -4 \end{bmatrix}^{-15}, x_8 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}^{-24}$$

$\mathrm{cov} = \dfrac{\sum (X_\mu - X_i)(Y_\mu - Y_i)}{N-1}$

(a) Calculate the covariance matrix of X. (I suggest you use a calculator!) $\begin{bmatrix} 13.1429 & -7.4286 \\ -7.4286 & 4.8571 \end{bmatrix}$

(b) Explain the sequence of steps involved in applying PCA (Principal Components Analysis) to a set of data, and how the result should be interpreted. $-$ mean / $\sqrt{\text{var}}$ UDU^T

(c) Apply PCA to the data set X. Write down the two Principal Components and the variance of X in the directions of each of the Principal Components. *(If you know how to calculate eigenvalues and eigenvectors then do this by hand. If you don't know how to calculate eigenvalues and eigenvectors but you've done it in the past then revise it and do it by hand. If you don't know how to calculate eigenvalues and eigenvectors and you've not done it in the past then use MATLAB (or similar).* $D \to \max_{1,2} \to U_1, U_2$

$A\vec{v} = \lambda\vec{v}$
$(A - \lambda I)\vec{v} = \vec{0}$
$\det(A - \lambda I) = 0$

4. A 6 dimensional data set has sample covariance matrix C, which has eigenvalue decomposition $C = UDU^T$

(i) What are the properties of the matrices D and U?

(ii) Given a 6-dimensional vector v, explain how you would calculate the projection of v onto the two most significant principal components of the data set.

D: diagonal with eigenvalues Stretching
U: orthorgonal $\det(U) = 1$ Rotation
$v \cdot x_1$
$v \cdot x_2$