LDA AND SVD

Assignment 1

ANU TS

January 12, 2023

Feature Reduction

Data may consist of many features. As the number of features increases, the system takes longer computation time. For faster computations, we must either remove unwanted features or reduce the dimensions of the features by combining variables. Two popular techniques for feature reduction are-LDA(Linear Discriminant Analysis) and SVD(Singular Value Decomposition).

LDA

Linear Discriminant analysis is one of the most popular dimensionality reduction techniques used for supervised classification problems in machine learning. It is also considered a pre-processing step for modeling differences in ML and applications of pattern classification.
There are mainly 5 steps for performing LDA

1. Find the ddimensional mean vectors.

2. Find the scatter matrices

3.Find the eigenvectors (e1,e2,...,ed) and their corresponding eigenvalues $\lambda 1, \lambda 2, \lambda 3, \lambda 4 ... \lambda d$

4. Arrange the resultant eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix W such that every column represents an eigenvector.

5. Use this eigenvector matrix to transform the samples onto the new subspace.
Example: Compute the Linear Discriminant projection for the following two dimensional data-set.
Samples for class 1 : X1=(x1,x2)=(4,2),(2,4),(2,3),(3,6),(4,4) and Sample for class 2 : X2=(x1,x2)=(9,10),(6,8)
STEP 1: Computing the d-dimensional mean vectors

$\mu_1 = 1/N_1 \sum_{x \epsilon w1} x$ =1/5 *[$\begin{bmatrix} 4 \\ 2 \end{bmatrix}$ + $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ + $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ + $\begin{bmatrix} 3 \\ 6 \end{bmatrix}$ + $\begin{bmatrix} 4 \\ 4 \end{bmatrix}$] = $\begin{bmatrix} 3 \\ 3.8 \end{bmatrix}$

$\mu_2 = 1/N_2 \sum_{x \epsilon w2} x$ =1/5 *[$\begin{bmatrix} 9 \\ 10 \end{bmatrix}$ + $\begin{bmatrix} 6 \\ 8 \end{bmatrix}$ + $\begin{bmatrix} 9 \\ 5 \end{bmatrix}$ + $\begin{bmatrix} 8 \\ 7 \end{bmatrix}$ + $\begin{bmatrix} 10 \\ 8 \end{bmatrix}$] = $\begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}$

class means are: $\boxed{\bar{\mu}_1 = \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \; and \; \bar{\mu}_2 = \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}}$

STEP 2 : Computing the Scatter Matrices

within-class scatter matrix S_W $is\ computed\ by\ the\ following\ equation\ S_W = \sum_{i=1}^{c} S_i$

where $S_i = \sum_{x\epsilon D_i}^{n} (x-\mu_i)-(x-\mu_i)^T$ Covariance matrix of the first class:

$S_i = \sum_{x\epsilon w1}^{n} (x-\mu_1)-(x-\mu_1)^T = \left[\begin{bmatrix}4\\2\end{bmatrix} - \begin{bmatrix}3\\3.8\end{bmatrix}\right]^2 + \left[\begin{bmatrix}2\\4\end{bmatrix} - \begin{bmatrix}3\\3.8\end{bmatrix}\right]^2 + \left[\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3\\3.8\end{bmatrix}\right]^2 + \left[\begin{bmatrix}3\\6\end{bmatrix} - \begin{bmatrix}3\\3.8\end{bmatrix}\right]^2 + \left[\begin{bmatrix}4\\4\end{bmatrix} - \begin{bmatrix}3\\3.8\end{bmatrix}\right]^2 = \begin{bmatrix}1 & -0.25\\ -0.25 & 2.2\end{bmatrix}$

$\boxed{\bar{c}ovariance\ matrix\ of\ the\ first\ class\ S_1\ =\ cov(X1)}$

$S_2 = \sum_{x\epsilon w2}^{n} (x-\mu_2)-(x-\mu_2)^T = \left[\begin{bmatrix}9\\10\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^2 + \left[\begin{bmatrix}6\\8\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^2 + \left[\begin{bmatrix}9\\5\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^2 + \left[\begin{bmatrix}8\\7\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^2 + \left[\begin{bmatrix}10\\8\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^2 = \begin{bmatrix}2.3 & -0.05\\ -0.05 & 3.3\end{bmatrix}$

$\boxed{\bar{c}ovariance\ matrix\ of\ the\ first\ class\ S_2\ =\ cov(X2)}$

$Within{-}class\ scatter\ matrix: S_W = \sum_{i=1}^{c} S_i = S_1 + S_2 = \begin{bmatrix}1 & -0.25\\ -0.25 & 2.2\end{bmatrix} + \begin{bmatrix}2.3 & -0.05\\ -0.05 & 3.3\end{bmatrix} = \begin{bmatrix}3.3 & -0.3\\ -0.3 & 5.5\end{bmatrix}$

$$\boxed{\bar{W}ithin - class\ scatter\ matrix: S_W\ = \begin{bmatrix}3.3 & -0.3\\ -0.3 & 5.5\end{bmatrix}}$$

$Now\ compute\ Between - class\ scatter\ matrix: S_B$

$$S_B = (\mu_1 - \mu_2) * (\mu_1 - \mu_2)^T = \left[\begin{bmatrix}3\\3.8\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right] * \left[\begin{bmatrix}3\\3.8\end{bmatrix} - \begin{bmatrix}8.4\\7.6\end{bmatrix}\right]^T$$

$$= \begin{bmatrix}-5.4\\-3.8\end{bmatrix} * \begin{bmatrix}-5.4 & -3.8\end{bmatrix} = \begin{bmatrix}29.16 & 20.52\\ 20.52 & 14.44\end{bmatrix}$$

$Solving\ the\ generalized\ eigenvalue\ problem\ for\ the\ matrix\ S_W^{-1} * S_B$

$The\ LDA\ projection\ is\ then\ obtained\ as\ the\ solution\ of\ the\ generalized\ eigen\ value\ problem$

$i.e,\ S_W^{-1} * S_B = \lambda_W$

$|S_W^{-1} * S_B - \lambda_W| = 0$

$$= \left| \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}^{-1} * \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix} - \lambda * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right|$$

$$= \left| \begin{bmatrix} 0.3045 & 0.0166 \\ 0.0166 & 0.1827 \end{bmatrix} * \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix} - \lambda * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$= \left| \begin{bmatrix} 9.2213 - \lambda & 6.483 \\ 4.2339 & 2.9794 - \lambda \end{bmatrix} \right| = 0$$

$$= (9.2213 - \lambda)(2.9794 - \lambda) - (6.483)(4.2339) = 0$$
$$= (\lambda)^2) - 12.2007 * \lambda = 0$$
$$i.e, \lambda_1 = 0 \; and \; \lambda_2 = 12.0027$$

$STEP\ 4 : Selecting\ linear\ discriminant\ for\ the\ new\ feature\ subspace$

$Hence, \lambda_1 = 0$

$$\begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = \lambda_1 * \begin{bmatrix} w1 \\ w2 \end{bmatrix} = \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = 0 * \begin{bmatrix} w1 \\ w2 \end{bmatrix}$$

$$and \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_2 = \lambda_2 * \begin{bmatrix} w1 \\ w2 \end{bmatrix}$$

$$= \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = 12.0027 * \begin{bmatrix} w1 \\ w2 \end{bmatrix}$$

$$Thus W_1 = \begin{bmatrix} -0.5755 \\ 0.8178 \end{bmatrix}$$

$$and \boxed{\bar{W}_2 = \begin{bmatrix} 0.9088 \\ 0.4173 \end{bmatrix} = W^*}$$

SVD

The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices. The SVD of mxn matrix A is given by the formula :

- U: *mxn* matrix of the orthonormal eigenvectors of AA^T.
- V^T: transpose of a *nxn* matrix containing the orthonormal eigenvectors of A^{T}A.
- W: a *nxn* diagonal matrix of the singular values which are the square roots of the eigenvalues of A^TA.

Consider the matrix A = $\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

In order to find the SVD of matrix A first find the eigenvalues of AA^T

AA^T = $\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}$ = $\begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$

Applying the characteristic equation for the above matrix is $W - \lambda I = 0$

$AA^T - \lambda I = 0$

$\lambda^2 - 34\lambda + 225 = 0$

= $(\lambda - 25)(\lambda - 9)$

so our singular values are: $\sigma_1 = 5\,; \sigma_2 = 3$ Now we find the right singular vectors i.e orthonormal set of eigenvectors of AA^T. The eigenvalues of AA^T are 25, 9, and 0, and since AA^T.is symmetric we know that the eigenvectors will be orthogonal.

For $\lambda = 25$,

$AA^T - 25I$= $\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix}$

which can be row-reduces to : $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

unit vector in the direction of it is:

v$_1$ $=$ $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

Similarly, for $\lambda = 9, v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$

For the 3rd eigenvector, we could use the property that it is perpendicular to v1 and v2 such that:

v$_1^T v_3 = 0$

$v_2^T v_3 = 0$

Solving the above equation to generate the third eigenvector

$v_3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ -a \\ -a/2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{-1}{3} \end{bmatrix}$ Now, we calculate U using the formula $u_i = \frac{1}{\sigma} A v_i$ and this gives U = $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$.

Hence, our final SVD equation becomes:

A =A = $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & \frac{-2}{3} & \frac{1}{3} \end{bmatrix}$