Tutorial on Component Analysis

Computing PCA
Computing LDA
Computing Laplacian Eigenmaps
Computing LPP
Principal Component Analysis

\[ W_0 = \arg \max_W \text{tr}[W^T S_t W] \]

s.t. \[ W^T W = I \]

\[ S_t = \frac{1}{N} \sum (x_i - \mu)(x_i - \mu)^T \]

solution \[ S_t W = W \Lambda \]

• We need to perform eigen-analysis of \( S_t \)

• Assuming we need \( d \) components we need computations of order \( O(dF^2) \) (if \( F \) is large this is quite demanding)
Principle Component Analysis

\[ S_t = XX^T \quad X = [x_1 - \mu, \ldots, x_N - \mu] \]

- Lemma 1: Assume \( B = XX^T \) and \( C = X^TX \)

\( \Rightarrow B \) and \( C \) have the same positive eigenvalues \( \Lambda \)

\( \Rightarrow \) assuming \( N < F \) then eigenvectors \( U \) of \( B \) and \( V \) of \( C \) are related as \( U = XV\Lambda^{-\frac{1}{2}} \)

Using Lemma 1 we can compute \( U \) in \( O(N^3) \)
Principal Component Analysis

\[ X^T X = V \Lambda V^T \]

\( V \) is a \( N \times (N - 1) \) matrix with columns the eigenvectors

\( \Lambda \) is a \((N - 1) \times (N - 1)\) diagonal matrix of eigenvalues

\[ U = XV\Lambda^{-\frac{1}{2}} \]

\( V^T V = I \) but \( VV^T \neq I \)

\[ U^T XX^T U = \Lambda^{-\frac{1}{2}} \quad V^T X^T XX^T XV \Lambda^{-\frac{1}{2}} \]

\[ = \Lambda^{-\frac{1}{2}} \quad V^T V \Lambda V^T V \Lambda V^T V \Lambda^{-\frac{1}{2}} = \Lambda \]

\[ I \quad I \quad I \]
Principal Component Analysis

• Step 1: Compute dot product matrix \( X^T X = [(x_i - \mu)^T (x_j - \mu)] \)

• Step 2: Perform eigenanalysis of \( X^T X = V \Lambda V^T \)

• Step 3: Compute eigenvectors \( U = X V \Lambda^{-\frac{1}{2}} \)

\[
U_d = [u_1, ..., u_d]
\]

• Step 4: Compute \( d \) features \( Y = U^T X \)
Whitening

Lets have a look at the covariance of $Y$

$$YY^T = U^T XX^T U = \Lambda$$

$$W = U \Lambda^{-\frac{1}{2}}$$
Linear Discriminant Analysis

\[ W_o = \arg \max_W \text{ tr}[W^T S_b W] \quad \text{s.t. } W^T S_w W = I \]

the eigenvectors of \( S_w^{-1} S_b \) that correspond to the largest eigenvalues

\[ S_w = \sum_{j=1}^{C} S_j = \sum_{j=1}^{C} \sum_{x_i \in c_j} (x_i - \mu(c_j))(x_i - \mu(c_j))^T \]

\[ \text{rank}(S_w) = \min(F, N - C) \]

\[ S_b = \sum_{j=1}^{C} N_{c_j} \mu(c_j) \mu(c_j)^T \]

\[ \text{rank}(S_b) = \min(F, C - 1) \]
How can we deal with the singularity of $S_w$

- Perform first PCA and reduce the dimensions to $N - C$ using $U$

- Solve LDA on the reduced space and get $Q$ ($Q$ has $C - 1$ columns)

- Total transform $W = UQ$ ($y = Q^T U^T x$)
Linear Discriminant Analysis

\[ X = \begin{bmatrix} x_1^1 & x_2^1 & x_3^1 & x_1^2 & x_2^2 \end{bmatrix} \]

\[ \mu(c_2) = \frac{1}{2} (x_1^2 + x_2^2) \]

\[ \mu(c_1) = \frac{1}{3} (x_1^1 + x_2^1 + x_3^1) \]

\[ E_1 = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \]

\[ E_2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \]
Matrix is idempotent
Linear Discriminant Analysis

\[ X \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} = \begin{bmatrix} \mu(c_1) & \mu(c_1) & \mu(c_1) & \mu(c_2) & \mu(c_2) \end{bmatrix} \]

\[ X \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} X^T \]

\[ = \begin{bmatrix} \mu(c_1) & \mu(c_1) & \mu(c_1) & \mu(c_2) & \mu(c_2) \end{bmatrix} \begin{bmatrix} \mu(c_1) \\ \mu(c_1) \\ \mu(c_1) \\ \mu(c_2) \\ \mu(c_2) \end{bmatrix} \]

\[ = 3\mu(c_1)\mu(c_1)^T + 2\mu(c_2)\mu(c_2)^T \]
Linear Discriminant Analysis

\[ S_b = XMMX^T = XMX^T \]

\[ E_j = \frac{1}{N_{c_j}} 11^T \]

\[ M = \begin{bmatrix}
E_1 & 0 & 0 & 0 & 0 \\
0 & E_2 & 0 & 0 & 0 \\
0 & 0 & E_3 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & 0 & E_c
\end{bmatrix} = \text{diag}\{E_1, \ldots, E_c\} \]
Linear Discriminant Analysis

\[ S_1 = \sum_{i=1}^{3} (x_i^1 - \mu(c_1))(x_i^1 - \mu(c_1))^T \]

\[ = \sum_{i=1}^{3} x_i^1 x_i^1^T - \mu(c_1)x_i^1^T - x_i^1 \mu(c_1)^T + \mu(c_1)\mu(c_1)^T \]

\[ = \sum_{i=1}^{3} x_i^1 x_i^1^T - 3\mu(c_1)\mu(c_1)^T \]

\[ S_w = S_1 + S_2 \]

\[ = \sum_{i=1}^{3} x_i^1 x_i^1^T + \sum_{i=1}^{2} x_i^2 x_i^2^T - (3\mu(c_1)\mu(c_1)^T + 2\mu(c_2)\mu(c_2)^T) \]
Linear Discriminant Analysis

\[ S_w = \sum_{i=1}^{3} x_i^1 x_i^1^T + \sum_{i=1}^{2} x_i^2 x_i^2^T - (3\mu(c_1)\mu(c_1)^T + 2\mu(c_2)\mu(c_2)^T) \]

\[ = XX^T - XMX^T = X(I - M)X^T \]

\[ S_t \quad S_b \]

\[ M \] is idempotent
so is idempotent \[ I - M \]

\[ S_t = S_w + S_b \]
Simultaneous Diagonalisation

\[ W_o = \arg\max_W \ \text{tr}[W^T S_b W] \ \text{s.t.} \ W^T S_w W = I \]

\[ \Rightarrow W_o = \arg\max_W \ \text{tr}[W^T XMMX^T W] \]
\[ \text{s.t.} \ W^T X(I - M)(I - M)X^T W = I \]
Simultaneous Diagonalisation

• Assume that $W = UQ$

What do we want?

$U$ to diagonalise $S_w = X(I - M)(I - M)X^T$

What does this mean?

$W^T X (I - M)(I - M)X^T W = I$

$Q^T U^T X (I - M)(I - M)X^T U Q = I$

$I$

Hence $U^T X (I - M)(I - M)X^T U = I$
Simultaneous Diagonalisation

\[ W_o = \arg\max_W \text{ tr}[W^T S_b W] \quad \text{s.t.} \quad W^T S_w W = I \]

\[ W = UQ \]

\[ \Rightarrow Q_o = \arg\max_W \text{ tr}[Q^T U^T X M M X^T U Q] \]

\[ \text{s.t.} \quad Q^T Q = I \]

Hence the constraint \[ W^T X (I - M) (I - M) X^T W = I \]

became \[ Q^T Q = I \]
Simultaneous Diagonalisation

(1) Find matrix $U$ such that

$$U^T X (I - M)(I - M)X^T U = I$$

$$X(I - M)(I - M)X^T = X_w X_w^T \quad X_w = X(I - M)$$

Lemma 1: We need to perform eigenanalysis to $X_w^T X_w$

$$X_w^T X_w = V_w \Lambda_w V_w^T \quad N - C \text{ positive eigenvalues}$$

$V_w$ is a $N \times (N - C)$ matrix

Hence $U = X_w V_w \Lambda_w^{-1}$
Simultaneous Diagonalisation

Can we verify that $U^T X (I - M)(I - M)X^T U = I$?

(2) Now we need to solve

$$Q_o = \arg \max_W \text{tr}[Q^T U^T X M M X^T U Q]$$

s.t. $Q^T Q = I$

$$\tilde{X}_b = U^T X M \quad \tilde{X}_b \text{ is a } (N - C) \times N \text{ matrix}$$

$$\Rightarrow \quad Q_o = \arg \max_W \text{tr}[Q^T \tilde{X}_b \tilde{X}_b^T Q] \quad \text{Does it ring a bell?}$$

s.t. $Q^T Q = I$
Simultaneous Diagonalisation

It is like doing PCA on the projected class means

\[ Q_o = \arg\max_W \text{tr}[Q^T \tilde{X}_b \tilde{X}_b^T Q] \]
\[ \text{s.t. } Q^T Q = I \]

\( Q_o \) is a matrix with columns the \( d \) eigenvectors \( \tilde{X}_b \tilde{X}_b^T \) that correspond to \( d \) largest eigenvalues \((d \leq C - 1)\)

\[ W_o = Q_o U \]
Simultaneous Diagonalisation

(1) Find the $p$ eigenvectors of $S_w$ that correspond to its non-zero eigenvectors (usually $N - C$)

$$U = [u_1, \ldots, u_{N-C}]$$

(2) Project the data $\tilde{X}_b = U^T XM$

(3) Perform PCA on $\tilde{X}_b$ to find $Q$

(4) Total transform is $W = UQ$
Laplacian Eigenmaps

\[
\min \operatorname{tr}[Y(D - S)Y^T] \quad \text{s.t.} \quad YDY^T = I
\]

(1) Find the $k$-nearest neighbours and construct matrix $S$

(make sure that $S$ is symmetric, $S = \frac{1}{2} (S + S^T)$).

(2) Compute the Laplacian $L = D - S$

(3) Perform eigenanalysis to $D^{-1} (D - S) = I - D^{-1} S$

and keep the eigenvectors that correspond to the smallest
Locality Preserving Projections

\[
\min_{Y = W^T X} \text{tr}[W^T X (D - S) X^T W] \quad \text{s.t.} \quad W^T X D X^T W = I
\]

Let’s do it on the board (it will be in the notes)