Tutorial on Component Analysis

Computing PCA
Computing LDA
Computing Laplacian Eigenmaps
Computing LPP

Principal Component Analysis

$$\boldsymbol{W}_0 = \arg\max_{\boldsymbol{W}} \operatorname{tr}[\boldsymbol{W}^T \boldsymbol{S}_t \boldsymbol{W}]$$

s.t. $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$ $\quad\quad$ $\boldsymbol{S}_t = \frac{1}{N}\sum (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$

solution $\boldsymbol{S}_t\boldsymbol{W} = \boldsymbol{W}\boldsymbol{\Lambda}$

- We need to perform eigen-analysis of $\boldsymbol{S}_t$
- Assuming we need d components we need computations of order $O(dF^2)$ (if F is large this is quite demanding)

Principal Component Analysis

$$\boldsymbol{S}_t = \boldsymbol{X}\boldsymbol{X}^T \qquad \boldsymbol{X} = [\boldsymbol{x}_1 - \boldsymbol{\mu}, \dots, \boldsymbol{x}_N - \boldsymbol{\mu}]$$

- Lemma 1: Assume $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{C} = \boldsymbol{X}^T\boldsymbol{X}$

$\Rightarrow$ $\boldsymbol{B}$ and $\boldsymbol{C}$ have the same positive eigenvalues $\boldsymbol{\Lambda}$

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

Using Lemma 1 we can compute $\boldsymbol{U}$ in $O(N^3)$

Principal Component Analysis

$$\boldsymbol{X}^T\boldsymbol{X} = \mathbf{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$$

$\mathbf{V}$ is a $Nx(N-1)$ matrix with columns the eigenvectors

$\boldsymbol{\Lambda}$ is a $(N-1)x(N-1)$ is a diagonal matrix of eigenvalues

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}$$

$\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$ but $\boldsymbol{V}\boldsymbol{V}^T \neq \boldsymbol{I}$

$$\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{U} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\ \boldsymbol{V}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}$$

$$= \boldsymbol{\Lambda}^{-\frac{1}{2}}\ \underbrace{\boldsymbol{V}^T\mathbf{V}}_{\boldsymbol{I}}\boldsymbol{\Lambda}\underbrace{\boldsymbol{V}^T\mathbf{V}}_{\boldsymbol{I}}\boldsymbol{\Lambda}\underbrace{\boldsymbol{V}^T\boldsymbol{V}}_{\boldsymbol{I}}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{\Lambda}$$

Principal Component Analysis

- Step 1: Compute dot product matrix $\boldsymbol{X}^T\boldsymbol{X} = [(\boldsymbol{x}_i - \boldsymbol{\mu})^T(\boldsymbol{x}_j - \boldsymbol{\mu})]$

- Step 2: Perform eigenanalysis of $\boldsymbol{X}^T\boldsymbol{X} = \mathbf{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$

- Step 3: Compute eigenvectors $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}$

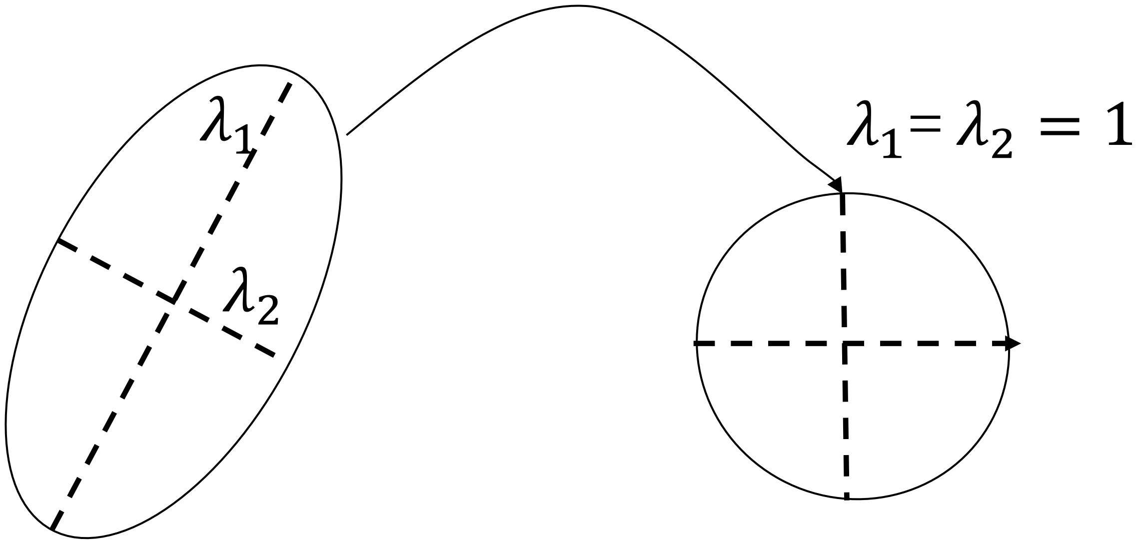
$$\boldsymbol{U}_d = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d]$$

- Step 4: Compute d features $\boldsymbol{Y} = \boldsymbol{U}^T\boldsymbol{X}$

Whitening

Lets have a look at the covariance of $\boldsymbol{Y}$

$$\boldsymbol{Y}\boldsymbol{Y}^{T} = \boldsymbol{U}^{T}\boldsymbol{X}\boldsymbol{X}^{T}\boldsymbol{U} = \boldsymbol{\Lambda}$$



$$\mathbf{W} = \boldsymbol{U}\boldsymbol{\Lambda}^{-\frac{1}{2}}$$

Linear Discriminant Analysis

$\boldsymbol{W}_o = \text{argmax}_{\boldsymbol{W}} \text{tr}[\mathbf{W}^{\mathrm{T}} \boldsymbol{S}_b \boldsymbol{W}]$ s.t. $\boldsymbol{W}^T \boldsymbol{S}_w \boldsymbol{W} = \mathbf{I}$

the eigenvectors of ${\boldsymbol{S}_w}^{-1} \boldsymbol{S}_b$ that correspond to the largest eigenvalues

$$\boldsymbol{S}_w = \sum_{j=1}^{C} \boldsymbol{S}_j = \sum_{j=1}^{C} \sum_{x_i \in c_j} (\boldsymbol{x}_i - \boldsymbol{\mu}(c_j))(\boldsymbol{x}_i - \boldsymbol{\mu}(c_j))^T$$

rank($\boldsymbol{S}_w$)=min$(F, N - C)$

$$\boldsymbol{S}_b = \sum_{j=1}^{C} N_{c_j} \boldsymbol{\mu}(c_j) \boldsymbol{\mu}(c_j)^{\mathrm{T}}$$

rank($\boldsymbol{S}_b$)=min$(F, C - 1)$

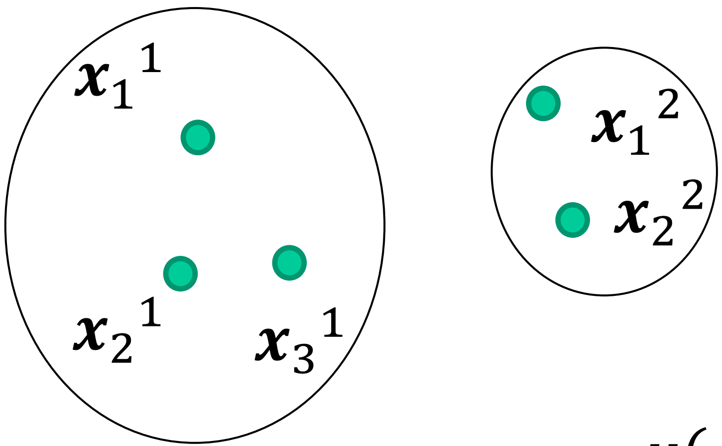
How can we deal with the singularity of $\boldsymbol{S}_w$

- Perform first PCA and reduce the dimensions to $N - C$ using $\boldsymbol{U}$

- Solve LDA on the reduced space and get $\boldsymbol{Q}$ ($\mathbf{Q}$ has $C - 1$ columns)

- Total transform $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{Q}$ ($\boldsymbol{y} = \boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{x}$)

Linear Discriminant Analysis



$$\boldsymbol{X} = [\underbrace{\boldsymbol{x}_1^{\,1} \;\; \boldsymbol{x}_2^{\,1} \;\; \boldsymbol{x}_3^{\,1}}_{c_1} \; \underbrace{\boldsymbol{x}_1^{\,2} \;\; \boldsymbol{x}_2^{\,2}}_{c_2}]$$

$$\boldsymbol{\mu}(c_2) = \frac{1}{2}(\boldsymbol{x}_1^{\,2} + \mathbf{x}_2^{\,2})$$

$$\boldsymbol{\mu}(c_1) = \frac{1}{3}(\boldsymbol{x}_1^{\,1} + \boldsymbol{x}_2^{\,1} + \boldsymbol{x}_3^{\,1})$$

$$\boldsymbol{E}_1 = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \qquad \boldsymbol{E}_2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

Linear Discriminant Analysis

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E}_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Matrix is idempotent

Linear Discriminant Analysis

$$\boldsymbol{X}\begin{bmatrix}\boldsymbol{E}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{E}_2\end{bmatrix} = [\boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_2)\ \boldsymbol{\mu}(c_2)]$$

$$\boldsymbol{X}\begin{bmatrix}\boldsymbol{E}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{E}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{E}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{E}_2\end{bmatrix}\boldsymbol{X}^T$$

$$= [\boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_1)\ \boldsymbol{\mu}(c_2)\ \boldsymbol{\mu}(c_2)]\begin{bmatrix}\boldsymbol{\mu}(c_1)\\ \boldsymbol{\mu}(c_1)\\ \boldsymbol{\mu}(c_1)\\ \boldsymbol{\mu}(c_2)^T\\ \boldsymbol{\mu}(c_2)\end{bmatrix}$$

$$= 3\boldsymbol{\mu}(c_1)\boldsymbol{\mu}(c_1)^T + 2\boldsymbol{\mu}(c_2)\boldsymbol{\mu}(c_2)^T$$

Linear Discriminant Analysis

$$\boldsymbol{S}_b = \boldsymbol{X}\boldsymbol{M}\boldsymbol{M}\boldsymbol{X}^T = \boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T$$

$$\boldsymbol{E_j} = \frac{1}{N_{c_j}}\boldsymbol{1}\boldsymbol{1}^{\boldsymbol{T}}$$

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{E}_3 & \boldsymbol{0} & \boldsymbol{0} \\ . & . & . & . & . \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{E}_c \end{bmatrix} = \text{diag}\{\boldsymbol{E_1}, \ldots, \boldsymbol{E_2}\}$$

Linear Discriminant Analysis

$$\boldsymbol{S}_1 = \sum_{i=1}^{3} ({\boldsymbol{x}_i}^1 - \boldsymbol{\mu}(c_1))({\boldsymbol{x}_i}^1 - \boldsymbol{\mu}(c_1))^T$$

$$= \sum_{i=1}^{3} {\boldsymbol{x}_i}^1 {{\boldsymbol{x}_i}^1}^T - \boldsymbol{\mu}(c_1){{\boldsymbol{x}_i}^1}^T - {\boldsymbol{x}_i}^1\boldsymbol{\mu}(c_1)^T + \boldsymbol{\mu}(c_1)\boldsymbol{\mu}(c_1)^T$$

$$= \sum_{i=1}^{3} {\boldsymbol{x}_i}^1 {{\boldsymbol{x}_i}^1}^T - 3\boldsymbol{\mu}(c_1)\boldsymbol{\mu}(c_1)^T$$

$$\boldsymbol{S}_w = \boldsymbol{S}_1 + \boldsymbol{S}_2$$

$$= \sum_{i=1}^{3} {\boldsymbol{x}_i}^1 {{\boldsymbol{x}_i}^1}^T + \sum_{i=1}^{2} {\boldsymbol{x}_i}^2 {{\boldsymbol{x}_i}^2}^T - (3\boldsymbol{\mu}(c_1)\boldsymbol{\mu}(c_1)^T + 2\boldsymbol{\mu}(c_2)\boldsymbol{\mu}(c_2)^T)$$

Linear Discriminant Analysis

$\boldsymbol{S}_w$

$$= \sum_{i=1}^{3} \boldsymbol{x}_i{}^1 \boldsymbol{x}_i{}^{1^T} + \sum_{i=1}^{2} \boldsymbol{x}_i{}^2 \boldsymbol{x}_i{}^{2^T} - (3\boldsymbol{\mu}(c_1)\boldsymbol{\mu}(c_1)^T + 2\boldsymbol{\mu}(c_2)\boldsymbol{\mu}(c_2)^T)$$

$$= \underbrace{\boldsymbol{X}\boldsymbol{X}^T}_{\boldsymbol{S}_t} - \underbrace{\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T}_{\boldsymbol{S}_b} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{X}^T$$

$\boldsymbol{M}$ is idempotent
so is idempotent $\boldsymbol{I} - \boldsymbol{M}$

$$\boldsymbol{S}_t = \boldsymbol{S}_w + \boldsymbol{S}_b$$

Simultaneous Diagonalisation

$$\boldsymbol{W}_o = \text{argmax}_{\boldsymbol{W}} \, \text{tr}[\mathbf{W}^{\mathrm{T}} \boldsymbol{S}_b \boldsymbol{W}] \;\; \text{s.t. } \boldsymbol{W}^T \boldsymbol{S}_w \boldsymbol{W} = \mathbf{I}$$

$$\Longrightarrow \boldsymbol{W}_o = \text{argmax}_{\boldsymbol{W}} \, \text{tr}[\boldsymbol{W}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{M}\boldsymbol{M}\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}]$$

$$\text{s.t. } \boldsymbol{W}^T \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{M})(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W} = \boldsymbol{I}$$

Simultaneous Diagonalisation

- Assume that $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{Q}$

What do we want?

$\boldsymbol{U}$ to diagonalise $\boldsymbol{S}_w = \boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T$

What does this mean?

$$\boldsymbol{W}^T\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T\boldsymbol{W} = \boldsymbol{I}$$

$$\boldsymbol{Q}^T\underbrace{\boldsymbol{U}^T\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T\boldsymbol{U}}_{\boldsymbol{I}}\boldsymbol{Q} = \boldsymbol{I}$$

Hence $\boldsymbol{U}^T\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T\boldsymbol{U} = \boldsymbol{I}$

Simultaneous Diagonalisation

$\boldsymbol{W}_o = \text{argmax}_{\boldsymbol{W}} \text{ tr}[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{W}]$ s.t. $\boldsymbol{W}^T\boldsymbol{S}_w\boldsymbol{W}=\mathbf{I}$

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{Q}$$

$$\Rightarrow \boldsymbol{Q}_o = \text{argmax}_{\boldsymbol{W}} \text{ tr}[\boldsymbol{Q}^T\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{U}\boldsymbol{Q}]$$

s.t. $\boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$

Hence the constraint $\boldsymbol{W}^T\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W} = \boldsymbol{I}$

became $\boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$

Simultaneous Diagonalisation

(1) Find matrix $\boldsymbol{U}$ such that

$$\boldsymbol{U}^T\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T\boldsymbol{U}=\boldsymbol{I}$$

$$\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}^T=\boldsymbol{X}_w{\boldsymbol{X}_w}^T \qquad \boldsymbol{X}_w=\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M})$$

Lemma 1: We need to perform eigenanalysis to ${\boldsymbol{X}_w}^T\boldsymbol{X}_w$

$${\boldsymbol{X}_w}^T\boldsymbol{X}_w=\boldsymbol{V}_w\boldsymbol{\Lambda}_w{\boldsymbol{V}_w}^T$$

$N-C$ positive eigenvalues

$\boldsymbol{V}_w$ is a $Nx(N-C)$ matrix

Hence $\boldsymbol{U}=\boldsymbol{X}_w\boldsymbol{V}_w{\boldsymbol{\Lambda}_w}^{-1}$

Simultaneous Diagonalisation

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

(2) Now we need to solve

$$\boldsymbol{Q}_o = \text{argmax}_{\boldsymbol{W}} \operatorname{tr}[\boldsymbol{Q}^T\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{U}\boldsymbol{Q}]$$
$$\text{s.t. } \boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$$

$\widetilde{\boldsymbol{X}}_b = \boldsymbol{U}^T\ \boldsymbol{X}\boldsymbol{M}$ $\qquad$ $\widetilde{\boldsymbol{X}}_b$ is a $(N-C)xN$ matrix

$$\Rightarrow \quad \boldsymbol{Q}_o = \text{argmax}_{\boldsymbol{W}} \operatorname{tr}[\boldsymbol{Q}^T\widetilde{\boldsymbol{X}}_b\widetilde{\boldsymbol{X}}_b{}^T\boldsymbol{Q}]$$
$$\text{s.t. } \boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$$

Does it ring a bell?

Simultaneous Diagonalisation

It is like doing PCA on the projected class means

$$\boldsymbol{Q}_o = \text{argmax}_{\boldsymbol{W}} \ \text{tr}[\boldsymbol{Q}^T \widetilde{\boldsymbol{X}}_b \widetilde{\boldsymbol{X}}_b{}^T \boldsymbol{Q}]$$
$$\text{s.t. } \boldsymbol{Q}^T \boldsymbol{Q} = \boldsymbol{I}$$

$\boldsymbol{Q}_o$ is a matrix with columns the d eigenvectors $\widetilde{\boldsymbol{X}}_b \widetilde{\boldsymbol{X}}_b{}^T$ that correspond to d largest eigenvalues $(d \leq C - 1)$

$$\boldsymbol{W}_o = \boldsymbol{Q}_o \boldsymbol{U}$$

Simultaneous Diagonalisation

(1) Find the p eigenvectors of $\boldsymbol{S}_w$ that correspond to its non-zero eigenvectors (usually $N - C$)

$$\boldsymbol{U} = [\boldsymbol{u}_{\boldsymbol{1}}, \dots, \boldsymbol{u}_{N-C}]$$

(2) Project the data $\widetilde{\boldsymbol{X}}_b = \boldsymbol{U}^T\ \boldsymbol{X}\boldsymbol{M}$

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

(4) Total transform is $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{Q}$

Laplacian Eigenmaps

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

(1) Find the k-nearest neighbours and construct matrix $\boldsymbol{S}$

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

(2) Compute the Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$

(3) Perform eigenanalysis to $\boldsymbol{D}^{-1}(\boldsymbol{D}-\boldsymbol{S}) = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{S}$
and keep the eigenvectors that correspond to the smallest

Locality Preserving Projections

$$\xrightarrow[\boldsymbol{Y}=\boldsymbol{W}^T\boldsymbol{X}]{} \min \operatorname{tr}[\boldsymbol{W}^T\boldsymbol{X}(\boldsymbol{D}-\boldsymbol{S})\boldsymbol{X}^T\boldsymbol{W}] \text{ s.t. } \boldsymbol{W}^T\mathbf{X}\boldsymbol{D}\boldsymbol{X}^T\boldsymbol{W}=\boldsymbol{I}$$

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