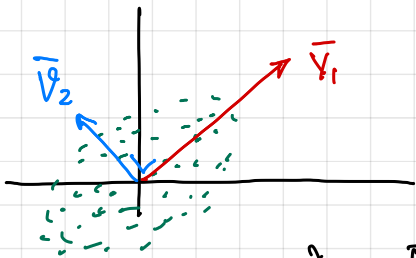
LECTURE 7A : PRINCIPLE COMPONENT ANALYSIS (PCA)

Also known as: Karhunen - Loeve Transform

Basic concept :



$\bar{v}_1$ = First principle component
= First eigen vector

$\bar{v}_2$ = Second eigen vector

λ_1 = First eigen value = largest variance

λ_2 = Second eigen value = second largest variance

Input : $\{\underset{D\times 1}{\bar{x}_n}\}_{n=1}^{N}$ = sample data

Output :

$\bar{v}_1 \ldots \bar{v}_M$ = the M eigenvectors
$\lambda_1 \ldots \lambda_M$ = the M largest eigenvalues

$\Big\}$ $M \leq D$

Q : How to get the eigenvectors & eigenvalues given the input ?

A : (1) Maximizing variance ; (2) Minimizing error

(1) Maximizing Variance :

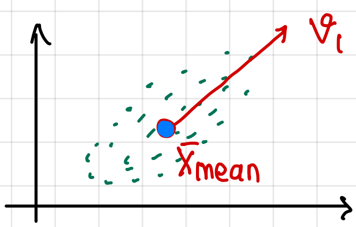
Suppose $\{\bar{x}_n\}_{n=1}^{N}$ are projected onto 1D space by $\bar{v}_1$ that maximizes the variance :

projection on $\bar{v}_1$ base vector

the origin of the base vectors

$$\boxed{var = \frac{1}{N}\sum_{n}^{N}\left(\underset{1\times D}{\bar{v}_1^T}\ \underset{D\times 1}{\bar{x}_n} - \underset{1\times D}{\bar{v}_1^T}\ \underset{D\times 1}{\bar{x}_{mean}}\right)^2}$$

variance of the samples after being projected onto a 1D space: $\bar{v}_1$



where : $\underset{D\times 1}{\bar{x}_{mean}} = \frac{1}{N}\sum_{n}^{N}\underset{D\times 1}{\bar{x}_n}$

$$\boxed{var = \frac{1}{N}\sum_{n}^{N}\Big(\underset{1\times D}{\bar{\vartheta}_1^T}\underset{D\times 1}{\bar{x}_n} - \underset{1\times D}{\bar{\vartheta}_1^T}\underset{D\times 1}{\bar{x}_{mean}}\Big)^2}$$

$$\left(\bar{\vartheta}_1^T\bar{x}_n - \bar{\vartheta}_1^T\bar{x}_{mean}\right)^2 = \Big[\underset{1\times D}{\bar{\vartheta}_1^T}\underset{D\times 1}{(\bar{x}_n - \bar{x}_{mean})}\Big]^2$$

Recall : $(ab)^2 = a^2 b^2$

Thus :
$$var = \frac{1}{N}\sum_{n}^{N}\left(\bar{\vartheta}_1^T\right)^2\left(\bar{x}_n - \bar{x}_{mean}\right)^2$$

$$= \frac{1}{N}\sum_{n}\underset{1\times D}{\bar{\vartheta}_1^T}\Big[\underset{D\times 1}{(\bar{x}_n - \bar{x}_{mean})}\underset{1\times D}{(\bar{x}_n - \bar{x}_{mean})^T}\Big]\underset{D\times 1}{\bar{\vartheta}_1}$$

$$= \bar{\vartheta}_1^T\Big[\frac{1}{N}\sum_{n}\underset{D\times 1}{(\bar{x}_n - \bar{x}_{mean})}\underset{1\times D}{(\bar{x}_n - \bar{x}_{mean})^T}\Big]\bar{\vartheta}_1$$

$$= \bar{\vartheta}_1^T\, S\, \bar{\vartheta}_1$$

where :
$$\boxed{S = \frac{1}{N}\sum_{n}(\bar{x}_n - \bar{x}_{mean})(\bar{x}_n - \bar{x}_{mean})^T}$$

Hence :
$$\max_{\{\bar{\vartheta}_1\}} var = \max_{\{\bar{\vartheta}_1\}} \bar{\vartheta}_1^T S\,\bar{\vartheta}_1$$
$$\text{s.t. } 1 - \bar{\vartheta}_1^T\bar{\vartheta} = 0$$

Without any constraint, the trivial solution is $|\bar{\vartheta}_1| = \infty$

Thus let's impose $\bar{\vartheta}_1^T\bar{\vartheta}_1 = 1$ ($\bar{\vartheta}_1$ is a unit vector)

$$\Downarrow$$

Using Lagrangian multiplier : $\max_{\{\bar{\vartheta}_1\}} L(\bar{\vartheta}_1) = \max_{\{\bar{\vartheta}_1\}} var$

where : $$\underset{1\times 1}{L(\bar{\vartheta}_1)} = \underset{1\times D}{\bar{\vartheta}_1^T}\underset{D\times D}{S}\underset{D\times 1}{\bar{\vartheta}_1} + \lambda_1\big(1 - \underset{1\times 1}{\bar{\vartheta}_1^T\bar{\vartheta}}\big)$$

Maximizing $L(\bar{\vartheta}_1)$:
$$\underset{D\times 1}{\frac{\partial L(\bar{\vartheta}_1)}{\partial\bar{\vartheta}_1}} = 2\underset{D\times D}{S}\underset{D\times 1}{\bar{\vartheta}_1} - 2\underset{D\times 1}{\lambda_1\bar{\vartheta}_1} = \underset{D\times 1}{0}$$

$$\Longleftrightarrow \boxed{S\,\bar{\vartheta}_1 = \lambda_1\bar{\vartheta}_1}$$

Therefore : $\left.\begin{array}{l}\lambda_1 = \text{the first eigen value}\\ \bar{\vartheta}_1 = \text{the first eigen vectors}\end{array}\right\}$ of S.

Q: Practically how can we get all the eigen vectors & eigenvalues of $\mathbb{S}$?

A: Use SVD (Singular Value Decomposition)

For our case, we want to solve:

$$\underset{D\times D}{\mathbb{S}}\ \underset{D\times 1}{\vec{v}_1} = \underset{1\times 1}{\lambda_1}\ \underset{D\times 1}{\vec{v}_1}$$: to find λ_1 & $\vec{v}_1$ given $\mathbb{S}$

if we want to have $\lambda_1 \dots \lambda_M$ and $\vec{v}_1 \dots \vec{v}_M$, then:

$$\underset{D\times D}{\mathbb{S}}\ \underset{D\times D}{\mathbb{V}} = \underset{D\times D}{\mathbb{V}}\ \underset{D\times D}{\lambda}$$; $\underbrace{\left[\vec{v}_1\ \vec{v}_2 \dots \vec{v}_D\right]}_{\mathbb{V}} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_D \end{bmatrix}}_{\lambda}$

For $M = 1$: $$\underset{D\times D}{\mathbb{S}}\ \underset{D\times 1}{\mathbb{V}} = \underset{D\times 1}{\mathbb{V}}\ \underset{1\times 1}{\lambda}$$

$$\underset{D\times D}{\mathbb{S}} = \underset{D\times D}{\mathbb{U}}\ \underset{D\times D}{\mathbb{D}}\ \underset{D\times D}{\mathbb{W}^T}$$ (= applying SVD to $\mathbb{S}$) ; you can use a Matlab / Numpy function for SVD.

Then; the column of $\underset{D\times D}{\mathbb{W}}$ is the eigenvectors of $\underset{D\times D}{\mathbb{S}}$: $\underset{D\times 1}{\vec{v}_1}, \underset{D\times 1}{\vec{v}_2} \dots \underset{D\times 1}{\vec{v}_D}$

since for our case, $\mathbb{U} = \mathbb{W}$, then $\mathbb{V} = \mathbb{U}$

Q: How about the eigenvalues?

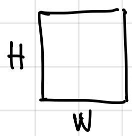
$$\mathbb{S}\mathbb{V} = \mathbb{V}\lambda \rightarrow \lambda = \mathbb{V}^{-1}\mathbb{S}\mathbb{V} = \mathbb{V}^T\mathbb{S}\mathbb{V}$$

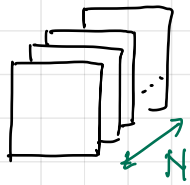
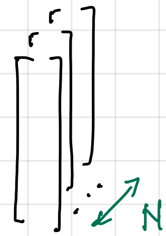
also: $$\mathbb{S} = \mathbb{U}\mathbb{D}\mathbb{W}^T \rightarrow \mathbb{U}^T\mathbb{S}\mathbb{W} = \mathbb{D}$$

$$\Big\} \ \lambda = \mathbb{D}$$

Reading: Pattern Recognition & Machine Learning (PRML), by Bishop
Chapter 12, Sect. 12.1

EIGENFACES

A face image:  $\rightarrow$ vectorization : $\underset{D\times 1}{\begin{bmatrix} \vdots \\ \vdots \end{bmatrix}}$ $D = W \cdot H$ $\qquad \underset{D\times 1}{\bar{x}_n} = \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$

N face images :  $\rightarrow$ Vectorization : 

1. Compute : $\underset{D\times 1}{\bar{x}_{mean}} = \frac{1}{N} \sum_{n=1}^{N} \underset{D\times 1}{\bar{x}_n}$

2. Form : S , where $\underset{D\times D}{S} = \frac{1}{N} \sum_{n} \underbrace{(\bar{x}_n - \bar{x}_{mean})}_{D\times 1} \underbrace{(\bar{x}_n - \bar{x}_{mean})^T}_{1\times D}$

3. Compute : $\underset{D\times D}{S} = \underset{D\times D}{U}\,\underset{D\times D}{D}\,\underset{D\times D}{V^T}$

the rows of $\underset{D\times D}{V^T}$ or the columns of $\underset{D\times D}{U}$ are the eigenvectors : $\bar{v}_1, \bar{v}_2 \dots \bar{v}_D$

4. Convert each of the eigenvectors $\bar{v}_1 \dots \bar{v}_D$ back to the image format $(H \times W)$

5. (optional) Face reconstruction :

$$\underset{D\times 1}{\tilde{x}_n} = \underset{D\times 1}{\bar{x}_{mean}} + \underbrace{\underset{D\times M}{U}\,\underset{M\times D}{U^T} \underbrace{(\bar{x}_n - \bar{x}_{mean})}_{D\times 1}}_{\lambda \,=\, \text{eigen values} \,=\, \text{variance}}$$

Dimensionality reduction:

$$\underset{D\times 1}{\tilde{x}_n} = \underset{D\times 1}{\bar{x}_{mean}} + \sum_{i=1}^{M} \underset{1\times D}{\bar{v}_i^T} \underbrace{(\bar{x}_n - \bar{x}_{mean})}_{D\times 1} \underset{D\times 1}{\bar{v}_i} \quad ; \quad M < D$$

$$= \bar{x}_{mean} + \sum_{i=1}^{M} \underset{1\times 1}{\lambda_i} \bar{v}_i \quad ; \quad \lambda_i = \bar{v}_i^T (\bar{x}_n - \bar{x}_{mean})$$

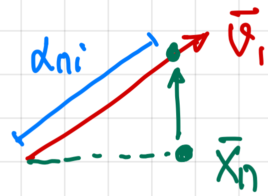
We can also represent $\underset{D\times 1}{\tilde{x}_n}$ using $\underset{M\times 1}{\hat{x}_n} = [\lambda_1, \lambda_2 \dots, \lambda_M]^T$

(2) Minimizing Error :

$$\underset{D\times 1}{\bar{x}_n} = \sum_{i=1}^{D} \alpha_{ni} \underset{D\times 1}{\vec{v}_i} \quad ; \quad \underbrace{\underset{1\times D}{\bar{v}_i^T} \underset{D\times 1}{\bar{v}_j} = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}}_{\text{orthogonal bases}}$$

Recall :

$$\underset{1\times 1}{\alpha_{ni}} = \underset{D\times 1}{\bar{x}_n} \cdot \underset{D\times 1}{\bar{v}_i} = \bar{x}_n^T \bar{v}_i$$



Hence :

$$\vec{x}_n = \sum_{i=1}^{D} (\bar{x}_n^T \bar{v}_i) \bar{v}_i$$

Our goal is to find $\tilde{x}_n$ that approximate $\bar{x}_n$ using M basis vectors (instead of D basis vectors) :

$$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} \bar{v}_i + \sum_{j=M+1}^{D} b_j \bar{v}_j \quad ;$$

b_j is constant for all data points/ samples.

Minimizing error :

$$J = \frac{1}{N} \sum_{n=1}^{N} |\bar{x}_n - \tilde{x}_n|^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left|\bar{x}_n - \sum_{i=1}^{M} z_{ni} \bar{v}_i + \sum_{j=M+1}^{D} b_j \bar{v}_j\right|^2$$

Setting $\frac{dJ}{\partial z_{ni}} = 0$: $\longrightarrow$ there are two indexes here : n & i

$$\underset{1\times 1}{\frac{\partial J}{\partial z_{ni}}} = \left(\bar{x}_n - \left[\sum_{i}^{M} z_{ni} \bar{v}_i + \sum_{j}^{D} b_j \bar{v}_j\right]\right) \bar{v}_i$$

$$\bar{x}_n^T \bar{v}_i - \sum_k z_{ni} \underbrace{\bar{v}_k^T \bar{v}_i}_{=1 \text{ if } k=i} - \sum_j b_j \underbrace{\bar{v}_j^T \bar{v}_i}_{j\neq i \;\; 0} = 0$$

$$\boxed{z_{ni} = \bar{x}_n^T \bar{v}_i}$$

Setting $\dfrac{\partial J}{\partial b_j} = 0$:

$\longrightarrow$ only one index : j

$$\underset{1\times1}{\frac{\partial J}{\partial b_j}} = \sum_{n=1}^{N} \left(\bar{x}_n - \left[\sum_{i=1}^{M} z_{ni}\,\bar{v}_i + \sum_{j=M+1}^{D} b_j \bar{v}_j\right]\right)\bar{v}_j = 0$$

$$= \sum_n \underset{1\times1}{\underline{\bar{x}_n^T \bar{v}_j}} - \sum_n \sum_{k=1}^{M} z_{nk} \underbrace{\bar{v}_k^T \bar{v}_j}_{=0}^{k\neq j} - \boxed{\sum_n \sum_{\ell=M+1}^{D} b_\ell\, \bar{v}_\ell^T \bar{v}_j} \quad \text{if } \ell = j : \rightarrow 1$$

$$= \sum_n \bar{x}_n^T \bar{v}_j - \sum_n b_j = 0$$

Hence :

$$\underset{1\times1}{b_j} = \frac{1}{N}\sum_{n=1}^{N} \underset{1\times1}{\underline{\bar{x}_n^T \bar{v}_j}}$$

$$\boxed{b_j = \bar{x}_{mean}^T\, \bar{v}_j}$$

$$\rightarrow \tilde{x}_n = \sum_{i=1}^{M} (\bar{x}_n^T \bar{v}_i)\,\bar{v}_i + \sum_{j=M+1}^{D} (\bar{x}_{mean}^T \bar{v}_j)\,\bar{v}_j$$

$$= \sum_{i=1}^{M} \lambda_i \bar{v}_i + \bar{x}_{mean}$$

Therefore :

$$J = \frac{1}{N}\sum_{n=1}^{N} |\bar{x}_n - \tilde{x}_n|^2$$

$$\bar{x}_n - \tilde{x}_n = \bar{x}_n - \sum_{i=1}^{M} \bar{x}_n^T \bar{v}_i\, \bar{v}_i - \sum_{j=M+1}^{D} \bar{x}_{mean}^T \bar{v}_j\, \bar{v}_j \quad ; \quad \bar{x}_n = \sum_{k=1}^{D} \overbrace{\bar{x}_n^T \bar{v}_k}^{=\lambda_k}\, \bar{v}_k$$

$$= \sum_{k=1}^{D} \bar{x}_n^T \bar{v}_k\, \bar{v}_k - \sum_{i=1}^{M} \bar{x}_n^T \bar{v}_i\, \bar{v}_i - \sum_{j=M+1}^{D} \bar{x}_{mean}^T \bar{v}_j\, \bar{v}_j$$

$$= \sum_{j=M+1}^{D} \bar{x}_n^T \bar{v}_j\, \bar{v}_j - \sum_{j=M+1}^{D} \bar{x}_{mean}^T \bar{v}_j\, \bar{v}_j$$

$$|\bar{x}_n - \tilde{x}_n|^2 = \left|\sum_j \left[(\bar{x}_n - \bar{x}_{mean})^T \bar{v}_j\right] \bar{v}_j\right|^2$$

Hence:

$$J = \frac{1}{N}\sum_n \sum_{j=M+1}^{D} \left[(\bar{x}_n - \bar{x}_{mean})^T \bar{v}_j\right]^2$$

$$= \sum_{j=M+1}^{D} \bar{v}_j^T \left(\frac{1}{N}\sum_n (\bar{x}_n - \bar{x}_{mean})(\bar{x}_n - \bar{x}_{mean})^T\right) \bar{v}_j = \sum_{j=M+1}^{D} \underset{1\times D}{\bar{v}_j^T}\, \underset{D\times D}{S}\, \underset{D\times 1}{\bar{v}_j}$$

Thus :

$$\boxed{\min_{\{\bar{v}_i\}} J = \min_{\{\bar{v}_i\}} \sum_{j=M+1}^{D} \bar{v}_j\, S\, \bar{v}_j \quad \text{s.t. } 1 - \bar{v}_j^T \bar{v}_j = 0}$$

PCA applications :

1. Data Compression

$$\underset{D\times 1}{\tilde{x}_n} = \sum_{i=1}^{M} z_{ni}\,\bar{v}_i + \sum_{j=M+1}^{D} b_j\,\bar{v}_j$$

$$= \sum_{i=1}^{M} (\underset{1\times D}{\bar{x}_n^T}\,\underset{D\times 1}{\bar{v}_i})\,\underset{D\times 1}{\bar{v}_i} + \sum_{i=M+1}^{D} (\underset{1\times D}{\bar{x}_{mean}^T}\,\underset{D\times 1}{\bar{v}_i})\,\underset{D\times 1}{\bar{v}_i}$$

$$\tilde{x}_n = \left[\sum_{i=1}^{M} (\underset{1\times D}{\bar{x}_n^T}\,\underset{D\times 1}{\bar{v}_i})\,\underset{D\times 1}{\bar{v}_i} + \sum_{i=M+1}^{D} (\underset{1\times D}{\bar{x}_{mean}^T}\,\underset{D\times 1}{\bar{v}_i})\,\underset{D\times 1}{\bar{v}_i}\right] + \underbrace{\left[\sum_{i=1}^{M}(\bar{x}_{mean}^T\bar{v}_i)\bar{v}_i - \sum_{i=1}^{M}(\bar{x}_{mean}^T\bar{v}_i)\bar{v}_i\right]}_{0}$$

$$\tilde{x}_n = \underbrace{\sum_{i=1}^{D} (\bar{x}_{mean}^T\bar{v}_i)\bar{v}_i} + \left[\sum_{i=1}^{M} (\bar{x}_n^T\bar{v}_i)\bar{v}_i - \sum_{i=1}^{M} (\bar{x}_{mean}^T\bar{v}_i)\bar{v}_i\right]$$

$\bar{x}_{mean}$ is the origin of the basis vectors
Hence, the sum of these projection = $\bar{x}_{mean}$

$$\tilde{x}_n = \bar{x}_{mean} + \underbrace{\sum_{i=1}^{M} (\underset{1\times D}{\bar{x}_n^T}\,\underset{D\times 1}{\bar{v}_i} - \underset{1\times D}{\bar{x}_{mean}^T}\,\underset{D\times 1}{\bar{v}_i})\,\underset{D\times 1}{\bar{v}_i}}_{\text{compressed data : why ?}} = \bar{x}_{mean} + \sum_{i=1}^{M} \lambda_i \bar{v}_i$$

Since $\lambda_i = \bar{x}_n^T\bar{v}_i - \bar{x}_{mean}^T\bar{v}_i$

We need $\underset{M\times 1}{\hat{x}_n}$, much smaller than $\underset{D\times 1}{\bar{x}_n}$

$$\hat{x}_n = \begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_M\end{bmatrix}$$

$$\underset{D\times 1}{\tilde{x}_n} = \underset{D\times 1}{\bar{x}_{mean}} + \sum_{i=1}^{M} \lambda_i\,\underset{D\times 1}{\bar{v}_i}$$

LECTURE 7B: NONNEGATIVE MATRIX FACTORIZATION

Recall: Every vector can be expressed as linear combination of basis vectors

$$\underbrace{\tilde{x}_n}_{D\times 1} = \underbrace{\bar{x}_{mean}}_{D\times 1} + \sum_{i=1}^{M} \alpha_i \underbrace{\bar{v}_i}_{D\times 1}$$

e.g.:

$$\begin{bmatrix} 2 \\ 6 \end{bmatrix} = 2 \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{base 1}} + 6 \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\text{base 2}} = 2 \underbrace{\begin{bmatrix} 1 \\ 3 \end{bmatrix}}_{\text{base 1}} + 0 \underbrace{\begin{bmatrix} -6 \\ 2 \end{bmatrix}}_{\text{base 2}}$$

- if we have a 2D image, we can transform it to a vector.
- Eigenface is formed by some basis vectors.
- Using PCA, basis vectors can have negative values, which doesn't make sense in some applications, e.g. face representations.

Solution: Non-negative Matrix Factorization:

$$\min_{\mathbb{B}, \lambda} \| \underbrace{\mathbb{X}}_{D\times N} - \underbrace{\mathbb{B}}_{D\times M} \underbrace{\lambda}_{M\times N} \|_F^2$$

s.t. $\mathbb{B}_{ij} \geq 0$

$\lambda_{ij} \geq 0$

; $\underbrace{\mathbb{X}}_{D\times N} = [\underbrace{\bar{x}_1}_{D\times 1} \; \bar{x}_2 \; \dots \; \underbrace{\bar{x}_N}_{D\times 1}]$

$\mathbb{B} = \underbrace{[\underbrace{\bar{b}_1}_{D\times 1} \; \bar{b}_2 \; \dots \; \underbrace{\bar{b}_M}_{D\times 1}]}_{M = \#\text{ basis vectors}}$ = basis vectors

$\lambda = [\underbrace{\bar{\lambda}_1}_{M\times 1} \; \bar{\lambda}_2 \; \dots \; \underbrace{\bar{\lambda}_N}_{M\times 1}]$ = coefficients

For every pixel:

$$\mathbb{X}_{ij} = (\mathbb{B}\lambda)_{ij} = \sum_{k=1}^{M} \mathbb{B}_{ik} \lambda_{kj}$$

$$\underbrace{J}_{1\times 1} = \| \underbrace{\mathbb{X}}_{D\times N} - \underbrace{\mathbb{B}\lambda}_{D\times N} \|_F^2$$

Optimization:

$$\left. \frac{\partial J}{\partial \mathbb{B}} = 0 \quad ; \quad \frac{\partial J}{\partial \lambda} = 0 \right\}$$ can't give closed-form solutions.

Hence : use Gradient Descent :

$$\underset{1\times 1}{\lambda_{kj}^{new}} = \underset{1\times 1}{\lambda_{kj}^{old}} - \underset{1\times 1}{\alpha_{kj}} \underset{1\times 1}{\frac{\partial J}{\partial \lambda_{kj}}}\Bigg|_{\lambda_{kj}^{old}}$$

Q : What is $\frac{\partial J}{\partial \lambda_{kj}}$?

A : $\underset{1\times 1}{J} = \| \underset{D\times N}{X} - \underset{D\times N}{B\lambda} \|_F^2 \rightarrow \underset{1\times 1}{\frac{\partial J}{\partial \lambda_{kj}}} = 2(X - B\lambda)B$

$$= 2\left([B^T X]_{kj} - [B^T B\lambda]_{kj}\right)$$

$$\lambda_{kj}^{new} = \lambda_{kj}^{old} - \alpha_{kj}\left([\underset{M\times D}{B^T}\underset{D\times N}{X}]_{kj} - [\underset{M\times M}{B^T B}\underset{M\times N}{\lambda}]_{kj}\right)$$

Similarly :

$$B_{ik}^{new} = B_{ik}^{old} - \alpha_{ik}\left([\underset{D\times N}{X}\underset{N\times M}{\lambda^T}]_{ik} - [\underset{D\times M}{B}\underset{M\times M}{\lambda\lambda^T}]_{ik}\right)$$

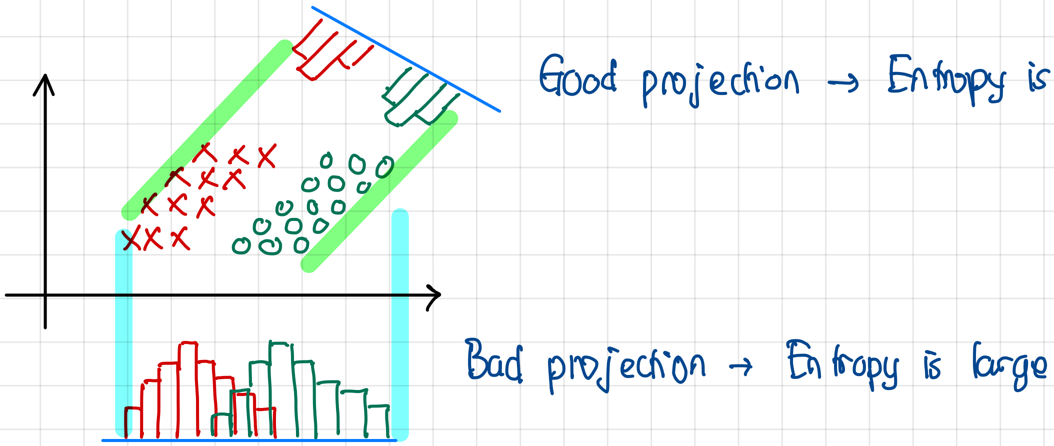
Considering Newton's method :

$$\alpha_{kj} = 1 \Big/ \frac{\partial^2 J}{\partial \lambda_{kj}^2} = \frac{1}{(B^T B)_{kj}} \approx \frac{\lambda_{kj}}{(B^T B\lambda)_{kj}}$$

Reading : Learning the parts of objects by non-negative matrix factorization, Lee & Seung , Nature 1999 .

LECTURE 7C : LINEAR DISCRIMINANT ANALYSIS (LDA)

Basic idea: To widen the separation between the projected means of different classes, and at the same time to narrow down the variance within each class.

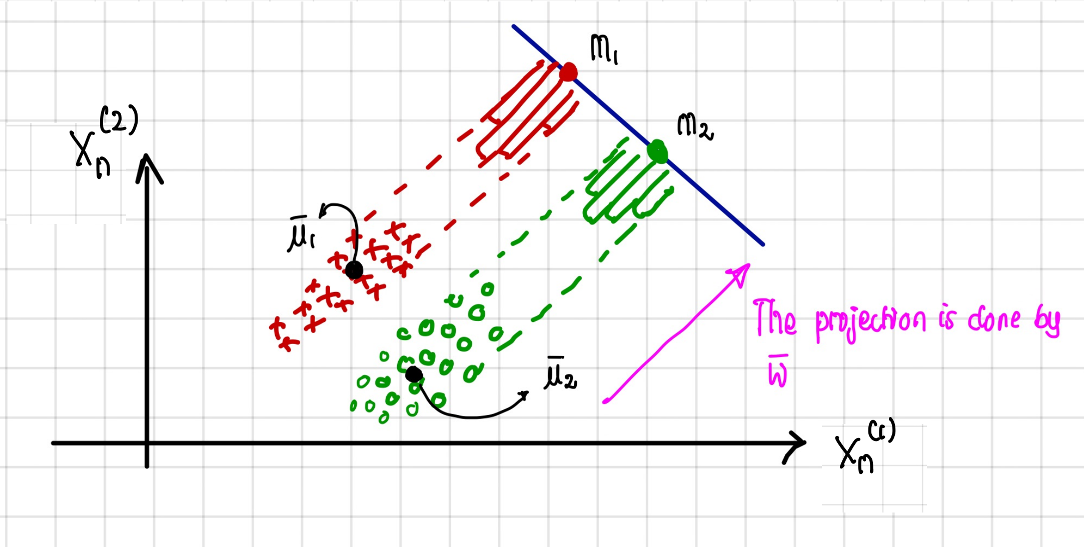


Consider 2-class classification :

① Mean & projected mean :

$$\underset{D\times 1}{\bar{\mu}_k} = \frac{1}{N} \sum_{n \in C_k} \underset{D\times 1}{\bar{x}_n} \quad ; \quad k = \{1,2\}$$

$$\underset{1\times 1}{m_k} = \underset{1\times D}{\bar{\omega}^T} \underset{D\times 1}{\bar{\mu}_k} \rightarrow \text{Projected mean by } \bar{\omega}$$



(2) Variance : $\begin{cases} \text{within class} \\ \text{between classes} \end{cases}$

- Variance within class :

$$\underset{1\times1}{S_k^2} = \sum_{n \in C_k} (\underset{1\times1}{y_n} - \underset{1\times1}{m_k})^2 \quad ; \quad y_n = \bar{w}^T \bar{x}_n$$

Total within class variance : $S_1^2 + S_2^2$

- Variance between classes : $\underset{1\times1}{B^2} = (\underset{1\times1}{m_2} - \underset{1\times1}{m_1})^2$

Fisher criteria :

$$J(\bar{w}) = \frac{B^2}{S_1^2 + S_2^2}$$

: To maximize :
maximizing variance between classes
& minimizing variance within class.

$$B^2 = (m_2 - m_1)^2 = [\underset{1\times D}{\bar{w}^T} (\underset{D\times1}{\bar{\mu}_2} - \underset{D\times1}{\bar{\mu}_1})]^2 = \underset{1\times D}{\bar{w}^T} [\underbrace{(\bar{\mu}_2 - \bar{\mu}_1)(\bar{\mu}_2 - \bar{\mu}_1)^T}_{D\times D}] \underset{D\times1}{\bar{w}}$$

$$B^2 = \underset{1\times D}{\bar{w}^T} \underset{D\times D}{S_B} \underset{D\times1}{\bar{w}}$$

S_B is known

$$S_1^2 + S_2^2 = \sum_{C_1} (y_n - m_1)^2 + \sum_{C_2} (y_n - m_2)^2$$

$$= \sum_{C_1} [\underset{1\times D}{\bar{w}^T} (\underset{D\times1}{\bar{x}_n} - \underset{D\times1}{\bar{\mu}_1})]^2 + \sum_{C_2} [\underset{1\times D}{\bar{w}^T} (\underbrace{\bar{x}_n - \bar{\mu}_2}_{D\times1})]^2$$

$$= \underset{1\times D}{\bar{w}^T} [\sum_{C_1} (\underbrace{\bar{x}_n - \bar{\mu}_1}_{D\times1})(\underbrace{\bar{x}_n - \bar{\mu}_1}_{1\times D})^T + \sum_{C_2} (\underbrace{\bar{x}_n - \bar{\mu}_2}_{D\times1})(\underbrace{\bar{x}_n - \bar{\mu}_2}_{1\times D})^T] \underset{D\times1}{\bar{w}}$$

$$S_1^2 + S_2^2 = \underset{1\times D}{\bar{w}^T} \underset{D\times D}{S_W} \underset{D\times1}{\bar{w}}$$

S_W is known

Hence :

$$\boxed{J(\bar{w}) = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}}}$$

: to maximize

Maximization: $\dfrac{\partial J(\bar{\omega})}{\underset{D\times 1}{\partial \bar{\omega}}} = \underset{D\times 1}{0}$; $J(\bar{\omega}) = \dfrac{\bar{\omega}^T S_B \bar{\omega}}{\bar{\omega}^T S_W \bar{\omega}}$

Recall:

$$\frac{\partial}{\partial x}\left(\frac{u}{v}\right) = \frac{v\,du - u\,dv}{v^2} \Rightarrow \begin{array}{l} u = \bar{\omega}^T S_B \bar{\omega} \\ v = \bar{\omega}^T S_W \bar{\omega} \end{array}$$

$$\underset{D\times 1}{\frac{\partial}{\partial \bar{\omega}} J(\bar{\omega})} = \frac{\underbrace{(\bar{\omega}^T S_W \bar{\omega})}_{1\times 1} \underbrace{S_B}_{D\times D} \underbrace{\bar{\omega}}_{D\times 1} - \underbrace{(\bar{\omega}^T S_B \bar{\omega})}_{1\times 1} \underbrace{S_W}_{D\times D} \underbrace{\bar{\omega}}_{D\times 1}}{\underbrace{(\bar{\omega}^T S_W \bar{\omega})^2}_{1\times 1}} = 0$$

$$= S_B \bar{\omega} - \frac{(\bar{\omega}^T S_B \bar{\omega})}{(\bar{\omega}^T S_W \bar{\omega})} S_W \bar{\omega} = 0$$

$$\Leftrightarrow \quad S_B \bar{\omega} = \frac{(\bar{\omega}^T S_B \bar{\omega})}{(\bar{\omega}^T S_W \bar{\omega})} S_W \bar{\omega}$$

$$\Leftrightarrow \quad \underbrace{(\bar{\omega}^T S_W \bar{\omega})}_{1\times 1} \underbrace{S_B}_{D\times D} \underbrace{\bar{\omega}}_{D\times 1} = \underbrace{(\bar{\omega}^T S_B \bar{\omega})}_{1\times 1} \underbrace{S_W}_{D\times D} \underbrace{\bar{\omega}}_{D\times 1}$$

$$c_1 S_B \bar{\omega} = c_2 S_W \bar{\omega} \qquad ; \quad c = c_1/c_2$$

$$\boxed{S_W \bar{\omega} = \frac{c_1}{c_2} S_B \bar{\omega}}$$

$$\bar{\omega} = c\, S_W^{-1} S_B \bar{\omega}$$

another constant

$$= c\, \underbrace{S_W^{-1}}_{D\times D} \underbrace{(\bar{\mu}_2 - \bar{\mu}_1)}_{D\times 1} \underbrace{(\bar{\mu}_2 - \bar{\mu}_1)^T}_{1\times D} \underbrace{\bar{\omega}}_{D\times 1}$$

projection of $(\mu_2 - \mu_1)$

$$\underbrace{\bar{\omega}}_{D\times 1} = c\, \underbrace{S_W^{-1}}_{D\times D} \underbrace{(\bar{\mu}_2 - \bar{\mu}_1)}_{D\times 1}$$

Hence:

$$\boxed{\bar{\omega} \propto S_W^{-1} (\bar{\mu}_2 - \bar{\mu}_1)}$$

Since we don't know c, we can only know the direction of $\bar{\omega}$.

$\bar{\omega}$ is the line separating the two classes.

S_W, $\bar{\mu}_2$, $\bar{\mu}_1$ are all known

MULTI-CLASSES LDA

Instead of 1 projection, we have C-1 projections (C = # classes)

$$\underset{1\times 1}{y_k} = \underset{1\times D}{\bar{w}_k^T} \underset{D\times 1}{\bar{x}_n} \quad \Rightarrow \quad \underset{K\times 1}{\bar{y}} = \underset{K\times D}{\bar{W}^T} \underset{D\times 1}{\bar{x}_n} \qquad ; \ K = C-1$$

Thus :

$$\underset{1\times 1}{J(W)} = \frac{\det(\underset{K\times D}{W^T} \underset{D\times D}{S_B} \underset{D\times K}{W})}{\det(\underset{K\times D}{W^T} \underset{D\times D}{S_W} \underset{D\times K}{W})}$$

$$W^* = \underset{\{W\}}{\operatorname{argmax}} \ J(W)$$

Solution : Recall from the previous page :

$$S_B \bar{w}_i = S_W \bar{w}_i \lambda_i \qquad ; \ \lambda_i = C$$

Thus :

$$\underset{D\times D}{S_B} \underset{D\times K}{W} = \underset{D\times D}{S_W} \underset{D\times K}{W} \underset{K\times K}{\lambda} \qquad M = H \times W \rightarrow \text{huge!}$$

$$\underbrace{S_W^{-1} S_B}_{\text{known}} \underset{D\times K}{W} = \underset{D\times K}{W} \underset{K\times K}{\lambda}$$

We can obtain W & λ, because : W is the eigenvectors of $S_W^{-1} S_B$
and λ is the eigen values of $S_W^{-1} S_B$

Fisherfaces : $\underset{M\times K}{W}$ are the K Fisher Faces.

Reading : Pattern Recognition & Machine Learning (PRML), by Bishop
Chapter 4, sect. 4.1