Dataset $X = [x_1 | x_2 | \dots | x_N]$ proj. vector u_1, $x_n, u_1 \in \mathbb{R}^\Delta$

Project X on u_1 => $\{u_1^T x_1, u_1^T x_2, \dots, u_1^T x_N\}$

Mean of projected samples is: $\frac{1}{N}\sum_{n=1}^{N} u_1^T x_n = u_1^T \underbrace{\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)}_{\bar{x}} = u_1^T \bar{x}$
(sample mean)

Sample variance of projected samples is: $\frac{1}{N}\sum_{n=1}^{N} (u_1^T x_n - u_1^T \bar{x})^2$

$Var(x) = E[(X - \underbrace{E[x]}_{\mu})^2]$

PCA opt. formulation for $M=1$:

$$u_1^* = \arg\max_{u_1} \frac{1}{N}\sum_{n=1}^{N} (u_1^T x_n - u_1^T \bar{x})^2$$

subj. to $u_1^T u_1 = 1$

$$\frac{1}{N}\sum_{n=1}^{N} \left(u_1^T (x_n - \bar{x})\right)^2$$

$$\frac{1}{N}\sum_{n=1}^{N} \underset{1\times\Delta}{u_1^T} \underset{\Delta\times 1}{(x_n - \bar{x})} \cdot \underset{1\times\Delta}{(x_n - \bar{x})^T} \underset{\Delta\times 1}{u_1}$$

$$u_1^T \underbrace{\left(\frac{1}{N}\sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T\right)}_{\Sigma} u_1$$

Σ the sample covariance of X

⇕

$$u_1^* = \arg\min_{u_1} -u_1^T \Sigma u_1$$

subj. to $u_1^T u_1 = 1$ $\quad u_1^T u_1 - 1 = 0$

$$\boxed{\begin{array}{l}\underset{u_1}{\operatorname{argmin}} \; -u_1^T \Sigma u_1 \\ \text{subj. to} \quad u_1^T u_1 - 1 = 0 \end{array}} \longrightarrow \lambda_1$$

, where $\Sigma = \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T$, $\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$

$\Sigma \longrightarrow \Delta \times \Delta$

The Lagrangian is $\mathcal{L}(u_1, \lambda_1) = -u_1^T \Sigma u_1 + \lambda_1 (u_1^T u_1 - 1)$

At the optimal solution for u_1, $\frac{\partial \mathcal{L}}{\partial u_1} = 0 \Rightarrow -u_1^T \underbrace{(\Sigma + \Sigma^T)}_{2\Sigma} + 2 \lambda_1 u_1^T = 0$

$\frac{\partial x^T C x}{\partial x} = x^T (C + C^T)$; $x \to u_1$, $C \to \Sigma$

$\Rightarrow 2 u_1^T \Sigma = 2 \lambda_1 u_1^T \Rightarrow u_1^T \Sigma = \lambda_1 u_1^T$

$\Rightarrow \Sigma u_1 = \lambda_1 u_1$

We want to max. sample variance $u_1^T \Sigma u_1$

$\Bigg\} \Rightarrow \max u_1^T (\lambda_1 u_1)$

$\Rightarrow \max \lambda_1 (u_1^T u_1) \Rightarrow \max \lambda_1 \Rightarrow \lambda_1$ is the largest (dominant) eigenvalue $\Rightarrow$ $\boxed{u_1 \text{ is the dominant eigenvector of } \Sigma}$ $M = 1$.

if Σ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq \dots \geq \lambda_\Delta$

$\qquad\qquad\qquad\qquad\;\; u_1 \quad\; u_2 \quad \dots \quad u_M \quad \dots \quad u_\Delta$

PCA for M, i.e. the M principal components are $u_1, u_2, \dots u_M$

Δ (u_j, λ_j) pairs that satisfy $\Sigma u_j = \lambda_j u_j$,

where $\{u_j\}_{1 \le j \le \Delta}$ form an orthonormal basis

$u_j^T u_i = 0, \ \forall i \neq j$

$\|u_j\| = 1, \ \forall j$

By virtue of being a sample covariance,

all $\lambda_j \geq 0$ for Σ.

λ_j are also the solution of the characteristic equation of Σ:

$$\det(\Sigma - \lambda I) = 0$$

import numpy.linalg as la

$[\Lambda, U] =$ la. eig (Σ)

$=$ la. eigh (Σ).

$U = [u_1 | u_2 | \ldots | u_\Delta]$

$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_\Delta)$

$$\Sigma u_j = \lambda_j u_j \quad \forall 1 \le j \le \Delta \iff \Sigma U = U \Lambda$$

$$U = [\,U_1 \mid U_2 \mid \cdots \mid U_\Delta\,] \Longrightarrow \Delta \times \Delta$$

U_1: $\Delta \times 1$

$$X \in \Delta \times 1$$

$$\underbrace{\underset{\Delta\times\Delta}{U^T}\ \underset{\Delta\times 1}{X}}_{\Delta \times 1} = \begin{bmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_\Delta^T \end{bmatrix} X = \underbrace{\begin{bmatrix} U_1^T X \\ U_2^T X \\ \vdots \\ U_\Delta^T X \end{bmatrix}}_{\Delta \times 1} = Y$$

$$U_{1,K} = \underbrace{[\,U_1 \mid U_2 \mid \cdots \mid U_K\,]}_{\Delta \times K}$$

$$\underbrace{U_{1,K}^T}_{K \times \Delta} X = \underbrace{\begin{bmatrix} u_1^T X \\ \vdots \\ U_K^T X \end{bmatrix}}_{K \times 1} = \hat{Y}$$

$$\hat{Y} = U_{1,K}^T X$$

$$\boxed{\hat{X} = \underbrace{U_{1,K}}_{\Delta \times K}\ \underbrace{U_{1,K}^T}_{K \times \Delta}\ \underbrace{X}_{\Delta \times 1} \in \mathbb{R}_{\Delta \times 1}}$$

$$\boxed{= U_{1,K}\, \hat{Y}}$$