Multi-variate Gaussian

① Linear Algebra

Part 0 PCA Recap

cover the
- a) Projection
- b) Basis transformation
- Coordinates tranformation

linear transformation

1) Maximize projection variance

本质: Problem Setting. $X \in \mathbb{R}^d \longrightarrow \mathbb{R}^m$

projection $\Rightarrow \underline{\begin{pmatrix} u_1^T \\ \vdots \\ u_m^T \end{pmatrix}} x \longrightarrow$ maximize variance

$$\Rightarrow X \cdot U := Z = \underline{\underline{\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}}}$$

$\downarrow$

maximize variance

2) Minimize Re-construct Error

Setting: $U_d \rightarrow (u_1, \dots, u_d) \longrightarrow$ orthogonal vectors

then linear transformation : $x \longmapsto Ux$

correspond to basis transformation

$(e_1, \dots, e_d) \longmapsto (u_1, \dots, u_d)$

for $x = (e_1, \dots, e_d) \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \rightarrow$ previous coordinates $= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$\downarrow$

$$x = (u_1, \dots, u_d) \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$$

$$= U_d \cdot \underline{U_d^T \cdot U_d} \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$$

$$= (e_1, \dots, e_d) \cdot U_d \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} \Rightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = U_d^T \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$\Rightarrow$ the coordinates after basis transformation

turns to $\underbrace{U_d^T \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}}$

then $X = (u_1, \dots, u_d) \cdot U_d^T \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$

$$= \sum_{i=1}^{d} (\underline{u_i^T x})\, u_i = (u_1, \dots, u_d) \begin{pmatrix} z_1 \\ \vdots \\ z_d \end{pmatrix}$$

the

we want $\tilde{x}_{km} = \sum_{i=1}^{m} \beta_{ki} \cdot u_i$

s.t $\boxed{\min \sum_{k=1}^{n} \| \tilde{x}_{km} - x_k \|_2^2}$

$\boxed{\text{Part 1}}$ $\rightarrow$ What is $\boxed{Q}^T x$ $\boxed{x \in \mathbb{R}^d}$

orthogonal

① $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \longrightarrow (e_1, \dots, e_d)$ 的投影(坐标)

② basis transformation :

$$(e_1, \dots, e_d) \xrightarrow{Q} (Q_1, \dots, Q_d)$$

$$\boxed{Q = (Q_1, \dots, Q_d) \in \mathbb{R}^{d \times d}}$$

then $x = Q \cdot Q^T x$

$$= (Q_1, \dots, Q_d)\, (Q^T x)$$

$$= \sum_{i=1}^{d} (Q^T x)_i \, Q_i$$

☆

$\Rightarrow$ $\boxed{Q^T x \in \mathbb{R}^d \text{ is the coordinates in } (Q_1, \dots, Q_d)}$

where $(\theta_1, \dots, \theta_d) \xleftarrow{Q} (e_1, \dots, e_d)$

Multi-variate Gaussian

① Discussion from 1-variate Gaussian

→ starting from 1-variate Gaussian

$\boxed{X \sim N(\mu, \sigma^2)}$

$$\Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

→ what will happen if we Standardize X ? $Z = \frac{X-\mu}{\sigma}$

$\left(\text{of course, } \begin{cases} \mathbb{E}[X] = 0 \\ Var[X] = 1 \end{cases} \right)$

Answer:

$$\Rightarrow \mathbb{P}(Z \leq z) = \mathbb{P}(X \leq \sigma z + \mu)$$

$$\Rightarrow \boxed{\begin{aligned} f_Z(z) &= \frac{d\mathbb{P}(Z \leq z)}{dz} \\ &= \sigma \cdot f_X(\sigma z + \mu) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\} \end{aligned}}$$

$$\Rightarrow \boxed{Z \sim N(0,1)}$$

② Generalize to Multi-variate cases

1) Consider $\begin{cases} X_1, \dots, X_n \sim N(0,1) \\ X_1, \dots, X_n \text{ mutually independent} \end{cases}$

$$P_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} x^T x\right)$$

$$\Rightarrow \boxed{X \sim \mathcal{N}(\underline{0}, I_n)}$$

2) Generally $\underline{X \sim \mathcal{N}(\mu, \Sigma)}$

$$\left[\text{from } \underline{\text{theorem}},\quad \exists B \in \mathbb{R}^{n\times n} \text{ invertible, s.t } \begin{array}{l} Z = B^{-1}(X-\mu) \\ Z \sim \mathcal{N}(\underline{0}, I_n) \end{array} \right]$$

$$\Rightarrow P_X(x_1, \dots, x_n) \overset{Z = B^{-1}(X-\mu)}{=\!=\!=} P_Z(z_1, \dots z_n) \cdot \left| \det\left(\frac{\partial Z}{\partial X}\right) \right|$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}(X-\mu)^T (BB^T)^{-1} (X-\mu) \right\} \cdot \left| \det(B^{-1}) \right|$$

$$= (2\pi)^{-\frac{n}{2}} \left| \det(BB^T) \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(X-\mu)^T (BB^T)^{-1} (X-\mu) \right\}$$

Moreover, since
$$\begin{aligned} \Sigma &= \mathbb{E}[(X-\mu)(X-\mu)^T] \\ &= \mathbb{E}[B Z Z^T B^T] \\ &= B\, \mathbb{E}[Z Z^T]\, B^T \\ &= BB^T \end{aligned}$$

$$\Rightarrow P_X(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} \left| \det(\Sigma) \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu) \right\}$$

③ $\boxed{\text{Interpretation}}$

if $P_X(x_1, \dots, x_n) \equiv C$ (contour graph)

then $(X-\mu)^T \Sigma^{-1} (X-\mu) = C'$ holds for some C'

$$\boxed{\text{Assume } |\Sigma| > 0}$$

then $\Sigma = Q \Lambda Q^T$, $\underline{\Lambda > 0}$ (since Σ symmetric , PD)

$\Rightarrow$ $(X-\mu)^T \Sigma^{-1} (X-\mu)$

$= (X-\mu)^T Q \Lambda^{-1} Q^T (X-\mu)$

$= \left[Q^T (X-\mu) \right]^T \Lambda^{-1} \left[Q^T (X-\mu) \right]$

$= \left[(Q \Lambda^{-\frac{1}{2}})^T (X-\mu) \right]^T \left[(Q \Lambda^{-\frac{1}{2}})^T (X-\mu) \right] = C'$

define $P = (Q \Lambda^{-\frac{1}{2}})^T (X-\mu)$ $\quad \mu \longrightarrow$ expectation vector

$\boxed{\text{Diagram}}$