Multivariate Normal Distributions

Qiang Liu
UT Austin

Multivariate Distributions

- Multivariate distributions, mean, covariance matrices.

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d$$

Probability density function (PDF)

$$p(x) = p([x_1 \cdots x_d])$$

① $p(x) \geq 0$

② $\int_{\mathbb{R}^d} p(x)\,dx = 1$

$$\mu = E_p[X] = \begin{bmatrix} E_p[x_1] \\ \vdots \\ E_p[x_d] \end{bmatrix}$$

$$E_p[x_1] = \int p(x)\, x_1\, dx$$

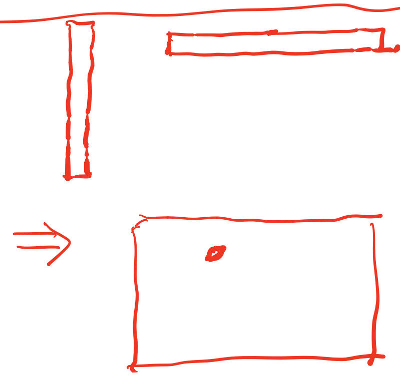
$$\boxed{\Sigma = [\mathrm{cov}(x_i, x_j)]_{ij}}$$

covariance matrix

$(d \times d)$

$$\mathrm{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)]$$

$$\boxed{\Sigma = E[(X-\mu)(X-\mu)^T]}$$



$\Rightarrow$

Normal Distributions

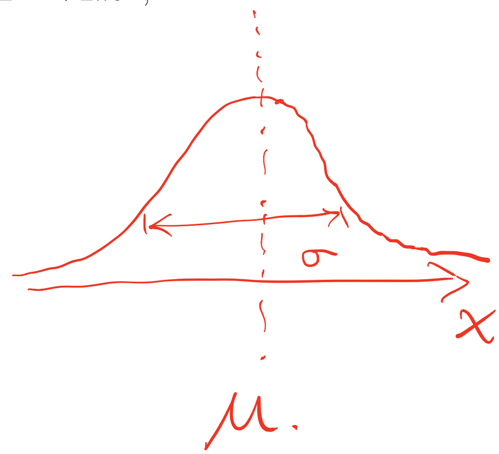
- A univariate distribution is called $\mathcal{N}(\mu, \sigma^2)$ if

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \qquad Z = \sqrt{2\pi\sigma^2},$$

Z

$$\mu = E_p[x] = \int x \, p(x) \, dx$$

$$\sigma^2 = E_p\left[(x-\mu)^2\right]$$



Given $\{x_i\}_{i=1}^n \sim \mathcal{N}(\mu, \sigma^2)$

Estimate μ, σ by:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2$$

Multivariate Normal Distributions

- A d-dimensional multivariate distribution is called $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ if

$$p(\boldsymbol{x}) = \frac{1}{Z}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad \boxed{Z = \sqrt{(2\pi)^d \det(\Sigma)}}$$

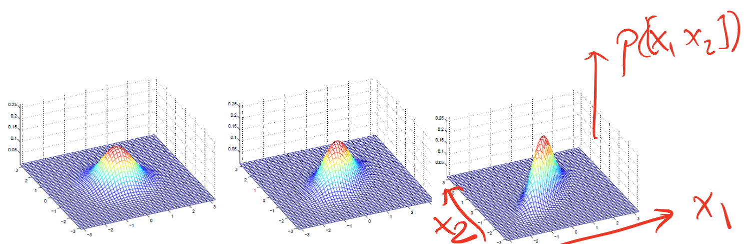
$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$\underbrace{\boxed{}}_{1\times d\ \to\ (x-\mu)^T} \ \underbrace{\boxed{\Sigma^{-1}}}_{d\times d} \ \underbrace{\boxed{}}_{d\times 1}\ (x-\mu)$$

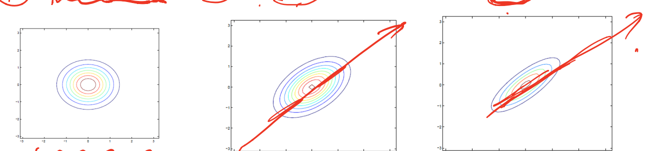
$$\sigma_{11} = E[(x_1-\mu_1)^2] = \sigma_1^2$$

$$\sigma_{12} = E[(x_1-\mu_1)(x_2-\mu_2)] = \mathrm{cov}(x_1, x_2) \geq 0$$

$$\Sigma = \begin{bmatrix} \sigma_1 & \sigma_{12} \\ \Sigma_1 & \sigma_{22} \end{bmatrix}$$

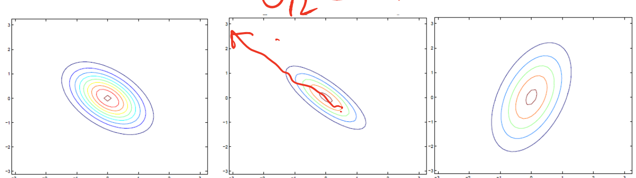


- μ = [0; 0]
- Σ = [1 0; 0 1]
- μ = [0; 0]
- Σ = [1 0.5; 0.5 1]
- μ = [0; 0]
- Σ = [1 0.8; 0.8 1]



$\sigma_{12} = 0.8$

$\sigma_{12} = -0.8$



- μ = [0; 0]
- Σ = [1 -0.5 ; -0.5 1]
- μ = [0; 0]
- Σ = [1 -0.8 ; -0.8 1]
- μ = [0; 0]
- Σ = [3 0.8 ; 0.8 1]

Figures from Pieter Abbeel's slides

Independent Normal Distributions

$\mu_i \quad \sigma_i^2$

$x_i \perp x_j \iff P(x_i, x_j) = P(x_i)P(x_j)$

- Let $x_i \sim \mathcal{N}(\mu_i,\ \sigma_i^2), \quad i = 1, \dots, d.$
- Assume $\{x_i\}$ are independent with each other ($x_i \perp x_j$ for $i \neq j$).
- Then their concatenation $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ follows a multivariate normal...

$$\mu = E[x] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix}$$

$$\Sigma = E\left[(x-\mu)(x-\mu)^T\right] = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}$$

Marginal Distributions

- Assume $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We can divide the vectors into two blocks:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Σ_{12}

- Then the marginal distribution of $\boldsymbol{x}_1$ is also Gaussian:

$$\boldsymbol{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}).$$

$x_1 \in \mathbb{R}^{d_1}$

$x_2 \in \mathbb{R}^{d_2}$

$d_1 + d_2 = d$

$\Sigma_{11} \in \mathbb{R}^{d_1 \times d_1}$

$\Sigma_{12} \in \mathbb{R}^{d_1 \times d_2}$

$\Sigma_{12} = \Sigma_{21}^T$

Conditional Distributions

- Assume $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We can divide the vectors into two blocks:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

- The conditional distribution $p(\boldsymbol{x}_1 \mid \boldsymbol{x}_2 = \boldsymbol{a})$ is also Gaussian:

$$\boldsymbol{x}_1 \mid (\boldsymbol{x}_2 = \boldsymbol{a}) \sim \mathcal{N}(\boldsymbol{\mu}_{1|2}, \ \boldsymbol{\Sigma}_{1|2})$$

where

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \mathbb{E}[\boldsymbol{x}_1 \mid \boldsymbol{x}_2 = \boldsymbol{a}] = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{a} - \mu_2) \\ \boldsymbol{\Sigma}_{1|2} &= \operatorname{cov}(\boldsymbol{x}_1 \mid \boldsymbol{x}_2 = \boldsymbol{a}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

$\boldsymbol{\Sigma}_{1|2}$ is called the Schur complement of $\boldsymbol{\Sigma}_{22}$ in $\boldsymbol{\Sigma}$.

Linear Transform

- Let $\boxed{z \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$.
- Applying linear transform:

$$\boxed{\boldsymbol{x} = A\boldsymbol{z} + b.}$$

Then $\boldsymbol{x}$ is also a multivariate normal...

$$\boxed{\begin{aligned} &x \sim \mathcal{N}(\mu_x, \Sigma_x) \\ &\mu_x = A\mu + b \\ &\Sigma_x = A\Sigma A^T \end{aligned}}$$

$d_1 \times d_1 \quad (d_1 \times d)(d \times d)(d \times d_1)$

$$x \sim \mathcal{N}(\mu_x, \Sigma_x)$$

$$\mu_x = E[x] = E[Az + b]$$

$$= A\,E[z] + b = A\mu + b$$

$$\Sigma_x = E[(x - \mu_x)(x - \mu_x)^T]$$

$$= E[(Az + b - A\mu - b)(Az + b - A\mu - b)^T]$$

$$= E[A(z - \mu)(A(z - \mu))^T]$$

$$= E[A(z - \mu)(z - \mu)^T A^T]$$

$$= A\,E[(z - \mu)(z - \mu)^T]A^T = A\Sigma A^T$$

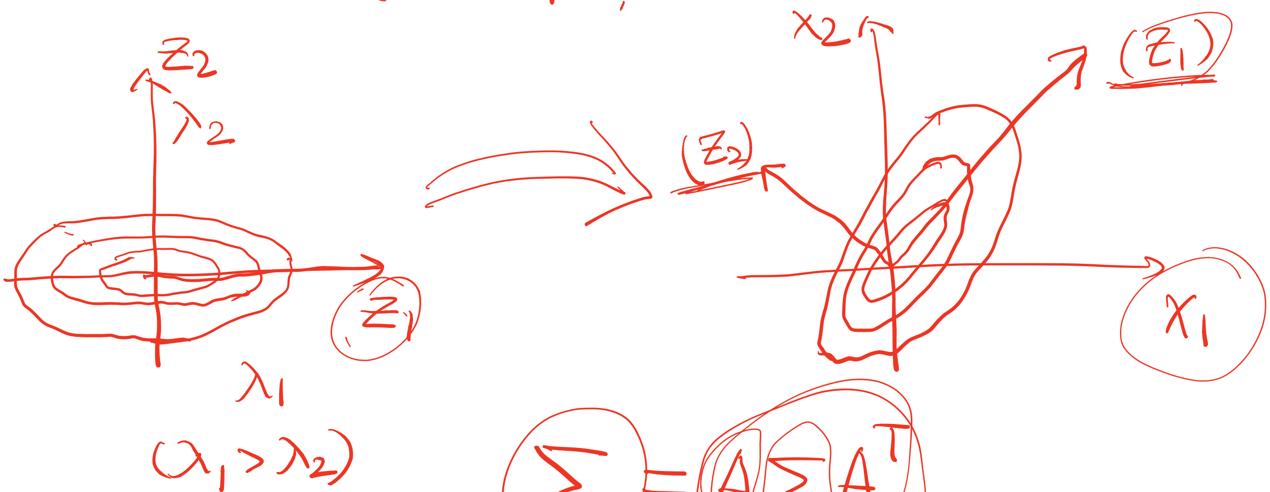
Multivariate Normal and PCA

- We can construct a probabilistic interpretation of principle component analysis (PCA) using multivariate normal distributions.

Assume $Z \sim \mathcal{N}(0, \Sigma)$, $\Sigma = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$

$X = AZ + b$

Observe $\{X^{(i)}\}_{i=1}^{n}$, Estimate Z ?



$\Sigma_x = A \Sigma A^T$

diagonal.

Maximum Likelihood Estimation

- Given an observation $\{\boldsymbol{x}_i\}_{i=1}^n$ drawn independently from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- We can show that the maximum likelihood estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ equals the empirical mean and covariance.

$$\max_{\mu, \Sigma} \sum_{i=1}^{n} \log p(x_i \mid \mu, \Sigma)$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right)\right)$$

$$= \underbrace{-n\log\left((2\pi)^{\frac{d}{2}}\right)}_{\text{const}} \boxed{-\frac{n}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$$

For $$L(Q) = \frac{n}{2}\log\det(Q) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\hat{\mu})^T Q (x_i-\hat{\mu})$$

$$\underbrace{\nabla_Q L(Q)}_{d \times d} = \frac{n}{2} Q^{-1} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = 0$$

$$\Rightarrow Q^{-1} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T.$$

Σ = Empirical Covariance.