Lecture3

Covariance

$$\begin{aligned} \mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}] \end{aligned}$$

$x \in \mathbb{R}^d$

$y \in \mathbb{R}^d$

$\mathrm{cov}[x, x] = E\left[\{x - E(x)\}\{x - E(x)\}^T\right]$

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathrm{cov}(x) = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{bmatrix}$

$$\bar{X} = [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_N]_{d \times N}$$

$$\bar{X} = \begin{bmatrix} x_{11} & x_{21} & & x_{N1} \\ x_{12} & x_{22} & & x_{N2} \\ x_{13} & x_{23} & & x_{N3} \end{bmatrix}_{d \times N} \leftarrow \text{data matrix}$$

$x_i \in \mathbb{R}^d \rightarrow i^{th}$ sample / data point / observation

$x_{i1} \in \mathbb{R} \rightarrow$ represents 1^{st} dimension of i^{th} sample

$$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$X = \begin{bmatrix} 1 & 3 & 4 & 7 \\ 2 & 2 & 0 & 1 \\ 0 & -1 & 5 & -1 \end{bmatrix}_{d\times N} \quad 3\times 4$$

$N=4,\ d=3$

$$\mathrm{cov}(x) = \frac{1}{N-1}\sum_{i=1}^{N}\underbrace{[x_i-\mu_x]}_{d\times 1}\underbrace{[x_i-\mu_x]^T}_{1\times d}$$

$$\mu_x = E[X] \in \mathbb{R}^{d\times 1} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\mu_x = \begin{matrix} 15/4 \\ 5/4 \\ 3/4 \end{matrix}$$

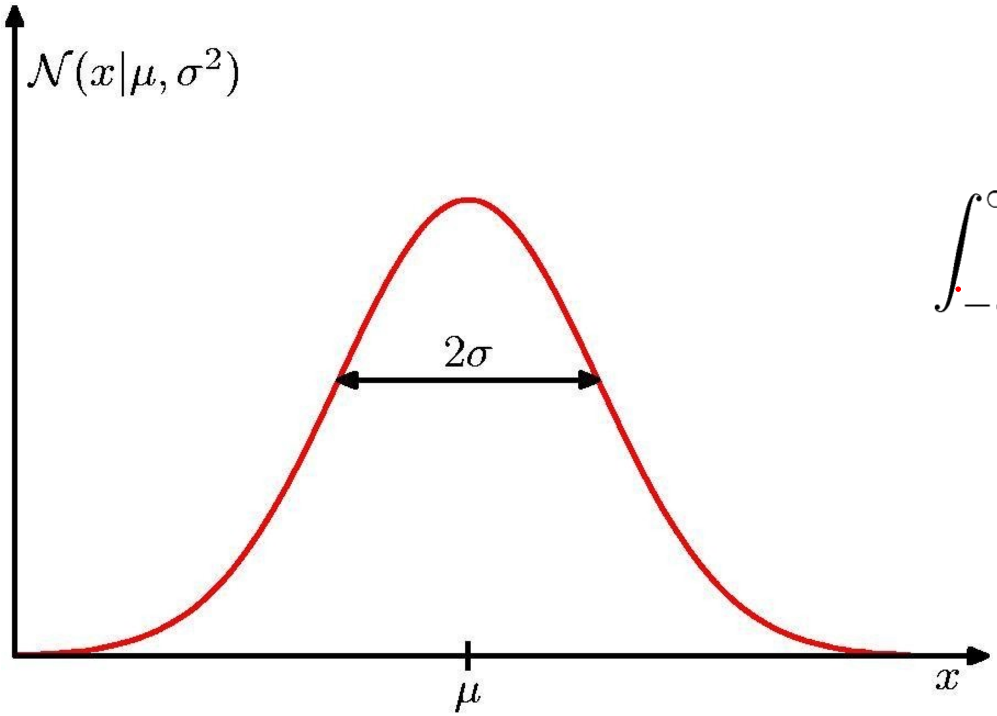
$$\mathrm{cov}(x) = \frac{1}{N-1}\left[(\bar{X}-\mu_x)(\bar{X}-\mu_x)^T\right]_{d \times d}$$

$$d \times N \qquad\qquad N \times d$$

x_1	$-$	μ_x
1		15/4
2	$-$	5/4
0		3/4

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

Gaussian Mean and Variance

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$[E(x)]^2 + E[(x-\mu)^2]$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

Ex.

- Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their sum satisfies

$E[x + z] = E[x] + E[z]$

$\mathrm{var}[x + z] = \mathrm{var}[x] + \mathrm{var}[z].$

$$\mu = E[x+z] = \int\int (x+z)\, p(x,z)\, \cancel{p(x)p(z)}\, dx\, dz$$

$$p(x,z) = p(x)\, p(z)$$

diff(a,b) = diff(b,a) $\Sigma \rightarrow$ Covariance of X

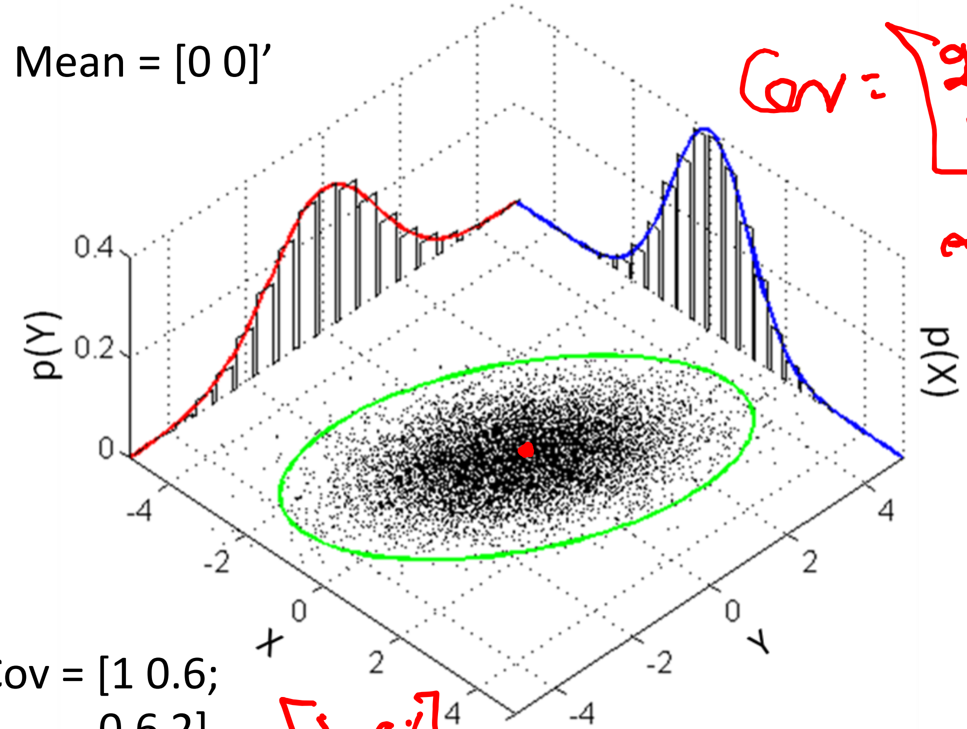
The Multivariate Gaussian

$X \in \mathbb{R}^d$

$D \rightarrow$ dimension

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} \underbrace{\overset{1\times d}{(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}} \overset{d\times d}{\boldsymbol{\Sigma}^{-1}} \overset{d\times 1}{(\mathbf{x}-\boldsymbol{\mu})}} \right\}$$

Mean = [0 0]'

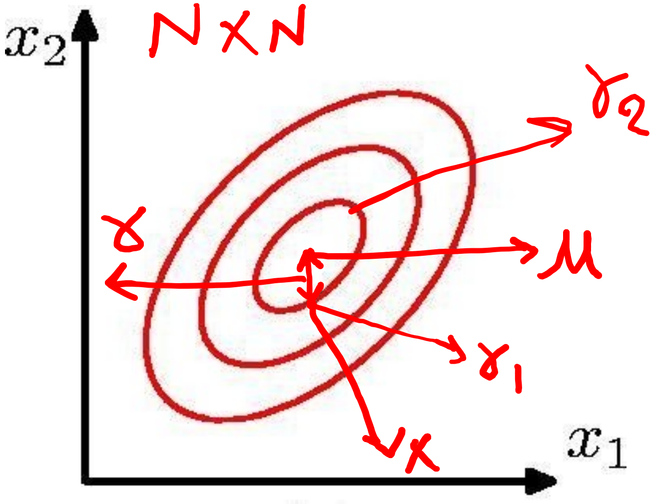


Cov = [1 0.6;
0.6 2]

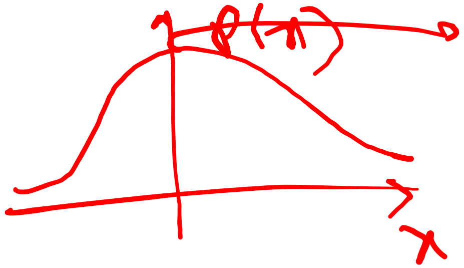
$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$

$\text{Cov} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow \sigma^2 I$

N×N

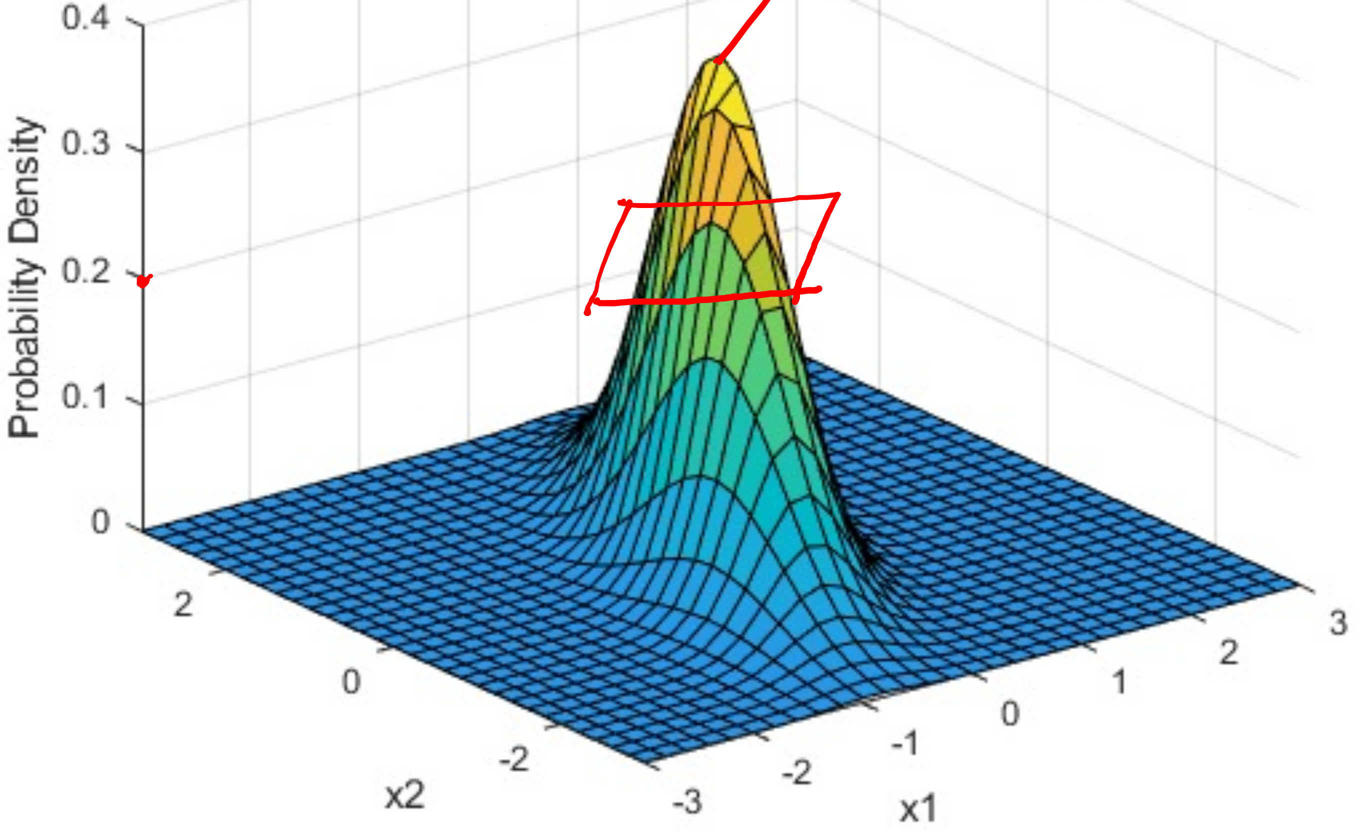


$\text{Cov} = \begin{bmatrix} \sigma^2_{x_1} & \sigma^2_{x_1,x_2} \\ \sigma^2_{x_2 x_1} & \sigma^2_{x_2} \end{bmatrix}$



mean

mean.



$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I \qquad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$P(X \mid \mu, \Sigma) = \frac{1}{(2\pi)^{2/2}\,\sigma^2} e^{\left\{ -\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{bmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right\}}$$

$$\mu = E(X) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$= \frac{1}{2\pi\sigma^2} e^{\left\{ -\frac{1}{2} \frac{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2}{\sigma^2} \right\}}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{ -\frac{1}{2} \frac{(x_1 - \mu_1)^2}{\sigma^2} \right\}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{ -\frac{1}{2} \frac{(x_2 - \mu_2)^2}{\sigma^2} \right\}}$$

$d = 100, \quad N = 10$

$N > d$

What will be max rank of cov matrix when d>N?

$\bar{X} \in \mathbb{R}^{d \times N}$

$\Sigma = \mathrm{Cov}(X) \in \mathbb{R}^{d \times d}$

$\max_{\text{rank}}(\Sigma) \leftarrow N$

$\Sigma \leftarrow$ not a full rank matrix.

What about $\Sigma^{-1/2}$?

$$\Sigma a = \lambda a$$

$$a^T \Sigma a = \lambda a^T a$$

Property $\rightarrow$ $\Sigma \rightarrow$ full rank. | symmetric.

$\rightarrow$ eigenvalues of $\Sigma > 0$

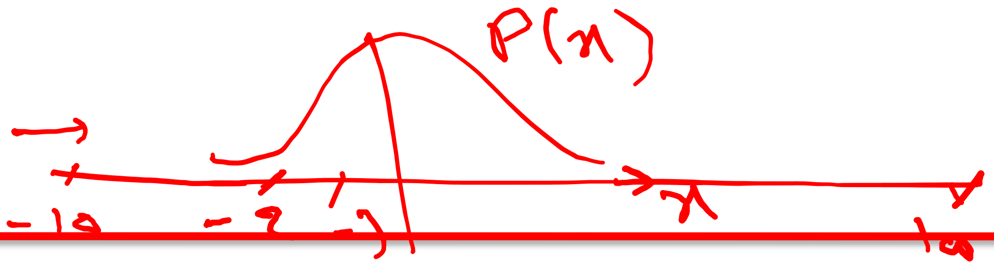
Positive Definite.

vector a, $a^T \Sigma a > 0$

$$a^T \left\{ \underbrace{\frac{1}{N-1} (\bar{X} - \mu)(\bar{X} - \mu)^T}_{} \right\} a > 0$$

$\Sigma \leftarrow$ not full rank, $\Sigma \rightarrow$ P.S.D.

$N(x, \mu, \sigma^2)$



$x \rightarrow$ r.v.

$\bar{x} = [\underbrace{-2, -3, -10, 100}_{x_1, x_2, x_2, x_3}]$

$P(x) \quad P(\bar{x})$

$P(x = x_1) \; P(x = x_2) \ldots P(x = x_n)$

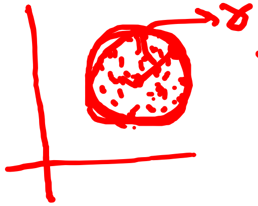
Squared Euclidean distance: $\underset{1\times d}{(X-\mu)^T} I \underset{d\times 1}{(X-\mu)} = \sum_{i=1}^{d} (x_i-\mu_i)^2$

Contd.

- $r^2 = (x - \mu)^t\Sigma^{-1}(x - \mu)$ is called the squared Mahalanobis distance from x to μ
- volume of the hyperellipsoid corresponding to a Mahalanobis distance r is given by

  $V = V_d|\Sigma|^{1/2}r^d$ where V_d is the volume of a d-dimensional unit hypersphere
- Higher the determinant for a fixed r and d, higher the scatter. For covariance matrices of independent variables, the determinant is large and thus the scatter is more.

$\Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$



$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Joint prob of N iid points of d-dimensions

$$X = \underbrace{[x_1 \quad x_2 \cdots\cdots x_N]}_{\to \text{joint prob}}$$

N-points. each point is d-dimensional

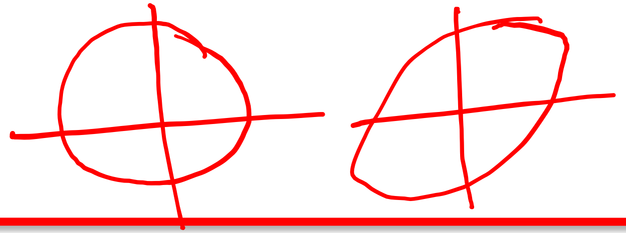
$x_i \in R^d$

i.i.d → identical independent distributed points.

$P(x_i) \leftarrow N(\mu, \Sigma)$ $\mu \in R^{d \times 1}$, $\Sigma \in R^{d \times d}$

$$p(x,y) = p(x)\,p(y)$$

$x, y \rightarrow$ scalar r.v. indep.



$$p(\bar{X}) = p(X_1, X_2 \cdots, X_N)$$

$$p(x_1) \leftarrow N(\mu, \Sigma)$$

$$p(x_2) \leftarrow N(\mu, \Sigma)$$

$$\ln p(\bar{X}) = \prod_{i=1}^{N} p(x_i) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} e^{\left\{-\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right\}}$$

$$= \sum_{i=1}^{N} \ln p(x_i)$$

i.i.d

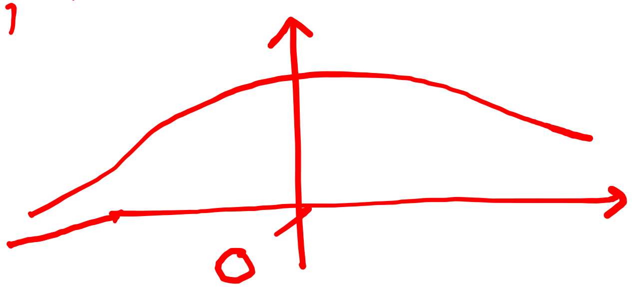
Two r.v. x & y , $x, y \in R$

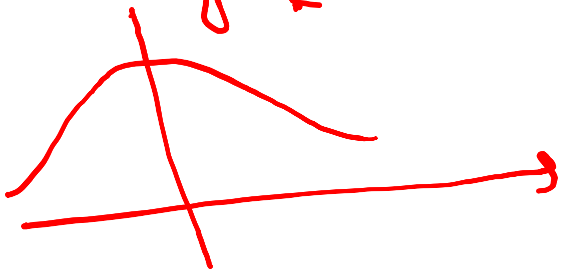
i.i.d → x & y are independent

but follow same distribution.

$(\mu, \sigma^2) \in p(x)$

$p(x) \to (\mu, \sigma^2)$





Chapter 2 (Part 1): Bayesian Decision Theory (Sections 2.1-2.2)

Introduction

Bayesian Decision Theory–Continuous Features

Introduction

Classes $\in \{\omega_1, \omega_2\}$

The sea bass/salmon example

State of nature, prior

classes

State of nature is a random variable

The catch of salmon and sea bass is equiprobable

$P(\omega_1) = P(\omega_2)$ (uniform priors)

$P(\omega_1) + P(\omega_2) = 1$ (exclusivity and exhaustivity)

$\rightarrow P(\omega_1) = 0.6$ $P(\omega_2) = 0.4$

salmon seabass

Question

You catch a fish. Tell which one is it?
Assumption: You cannot see the fish.

Question

You catch a fish. You can see it, may be eat too. Tell which one is it?