Module 2, Part 2: Random Vectors, Covariance, Multivariate Normal Distribution

TMA4268 Statistical Learning V2023

Daesoo Lee,

Department of Mathematical Sciences, NTNU

NTNU
Norwegian University of Science and Technology

19/01/2023

DA

How should I start a lecture to motivate students in college?

One way to start a lecture to motivate students in college is to begin with an engaging and thought-provoking question or statement related to the topic of the lecture. This can help to capture the students' attention and interest right away. Additionally, it can be helpful to provide a clear overview of what the students will learn during the lecture and how it relates to their overall coursework and future goals. Providing real-world examples and connections can also help to make the material more relevant and interesting to the students.

Quiz
(advised by ChatGPT)

Quiz

https://www.whichfaceisreal.com/

PLAY ABOUT METHODS LEARN PRESS CONTACT BOOK CALLING BS

Click on the person who is real.

Quiz

https://www.whichfaceisreal.com/

$Y \in \{0,1\}$; qualitative response

$$\hat{Y} = \hat{f}(X)$$

- X: image
- $\hat{Y}$: your guess
- $\hat{f}$: your brain(?)

$$\mathrm{MSE} = \mathrm{E}\left[\left(Y - \hat{Y}\right)^2\right]$$

MSE: mean squared error

Course Schedule

- Daesoo Lee
- a PhD student at the Department of Mathematical Sciences, NTNU
  - Main study area: Deep Learning

Calendar week	Module	Date	Weekday	Topic	Resp. Instructor
2	**Module 1**	09.01.23	Monday	Introduction	Steffi
		12.01.23	Thursday	R-course	Self-study
		12.01.23	Thursday	R-course	Self-study
3	**Module 2**	16.01.23	Monday	Statistical learning 1	Daesoo
		19.01.23	Thursday	Statistical learning 2	Daesoo
		19.01.23	Thursday	Statistical learning RecEx	Emma
4	**Module 3**	23.01.23	Monday	Linear regression 1	Daesoo
		26.01.23	Thursday	Linear regression 2	Daesoo
		26.01.23	Thursday	Linear regression RecEx	Kenneth
5	**Module 4**	30.01.23	Monday	Classification 1	Daesoo
		02.02.23	Thursday	Classification 2	Daesoo
		02.02.23	Thursday	Classification RecEx	Emma
6	**Module 5**	06.02.23	Monday	Resampling 1	Steffi
		09.02.23	Thursday	Resampling 2	Steffi / Emma (Rmd and ggplot)
		09.02.23	Thursday	Resampling RecEx	Kenneth
7		13.02.23	Monday	Compulsory Exercise 1	Emma and Kenneth
		16.02.23	Thursday	Compulsory Exercise 1	All
		16.02.23	Thursday	Compulsory Exercise 1	All
8	**Module 6**	20.02.23	Monday	Model Selection, Regularization 1	Steffi
		23.02.23	Thursday	Model Selection, Regularization 2	Steffi
		23.02.23	Thursday	Model Selection, Regularization RecEx	Daesoo
9	**Module 7**	27.02.23	Monday	Moving Beyond Linearity	Steffi

Overview

- Random vectors
- The covariance and correlation matrix
- The multivariate normal distribution

Random Vector

Random Vector

Recap

$$\boldsymbol{X} \longrightarrow \boxed{f} \longrightarrow \boldsymbol{Y}$$

$$n \text{ samples} \left\{ \overbrace{\begin{array}{|c|c|c|c|} \hline x_{11} & x_{12} & \cdots & x_{1p} \\ \hline x_{21} & x_{22} & \cdots & x_{2p} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline x_{n1} & x_{n2} & \cdots & x_{np} \\ \hline \end{array}}^{p \text{ covariates}} \right. \qquad \qquad \begin{array}{|c|} \hline y_1 \\ \hline y_2 \\ \hline \vdots \\ \hline y_n \\ \hline \end{array}$$

Recap

$$\boldsymbol{X} = \overbrace{\begin{array}{|c|c|c|c|} \hline x_{11} & \boxed{x_{12}} & \cdots & x_{1p} \\ \hline x_{21} & x_{22} & \cdots & x_{2p} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline x_{n1} & x_{n2} & \cdots & x_{np} \\ \hline \end{array}}^{p \text{ covariates}} \Bigg\} \; n \text{ samples}$$

$$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)$$

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ \vdots \\ x_{nj} \end{pmatrix}$$ Random Variable

$$\boldsymbol{Y} = \begin{array}{|c|} \hline y_1 \\ \hline y_2 \\ \hline \vdots \\ \hline y_n \\ \hline \end{array}$$

Random Vector

Random vector

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{bmatrix}$$

age	X_1
height	X_2
weight	X_3

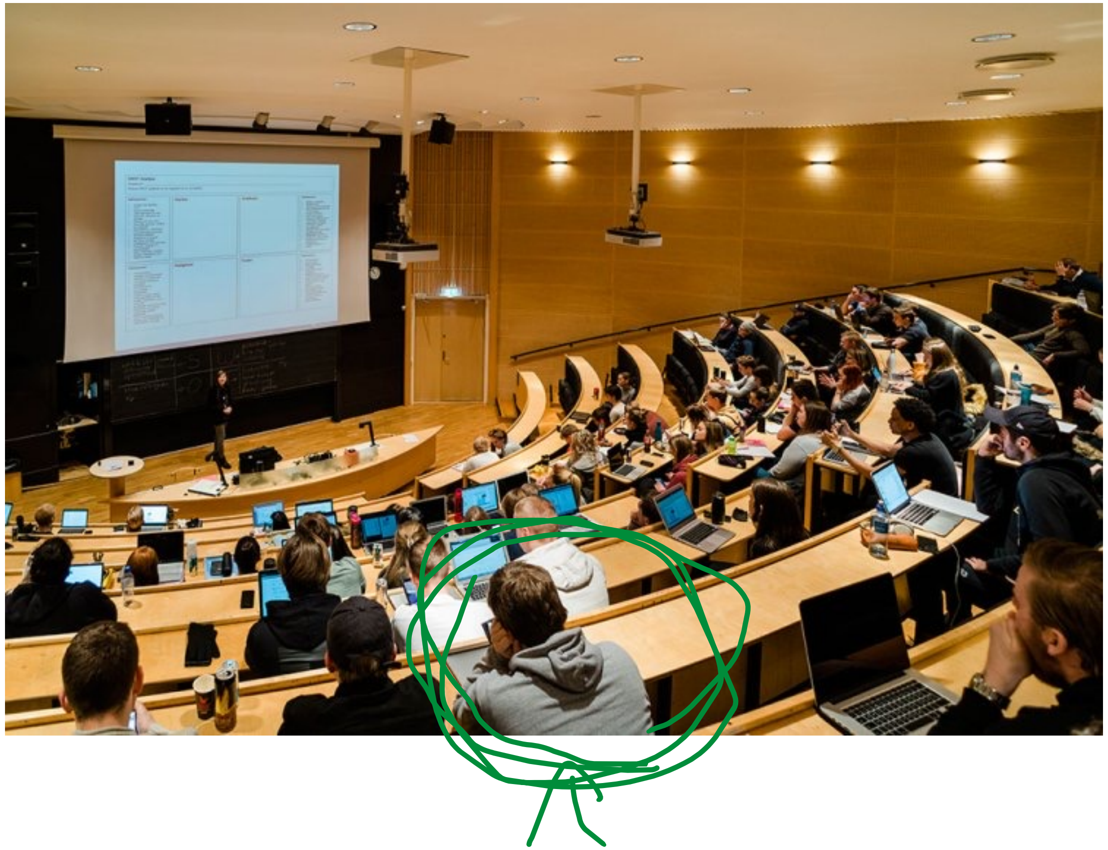
$$f(\boldsymbol{X}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$$

e.g., BMI

Random Vector

Random vector

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{bmatrix}$$

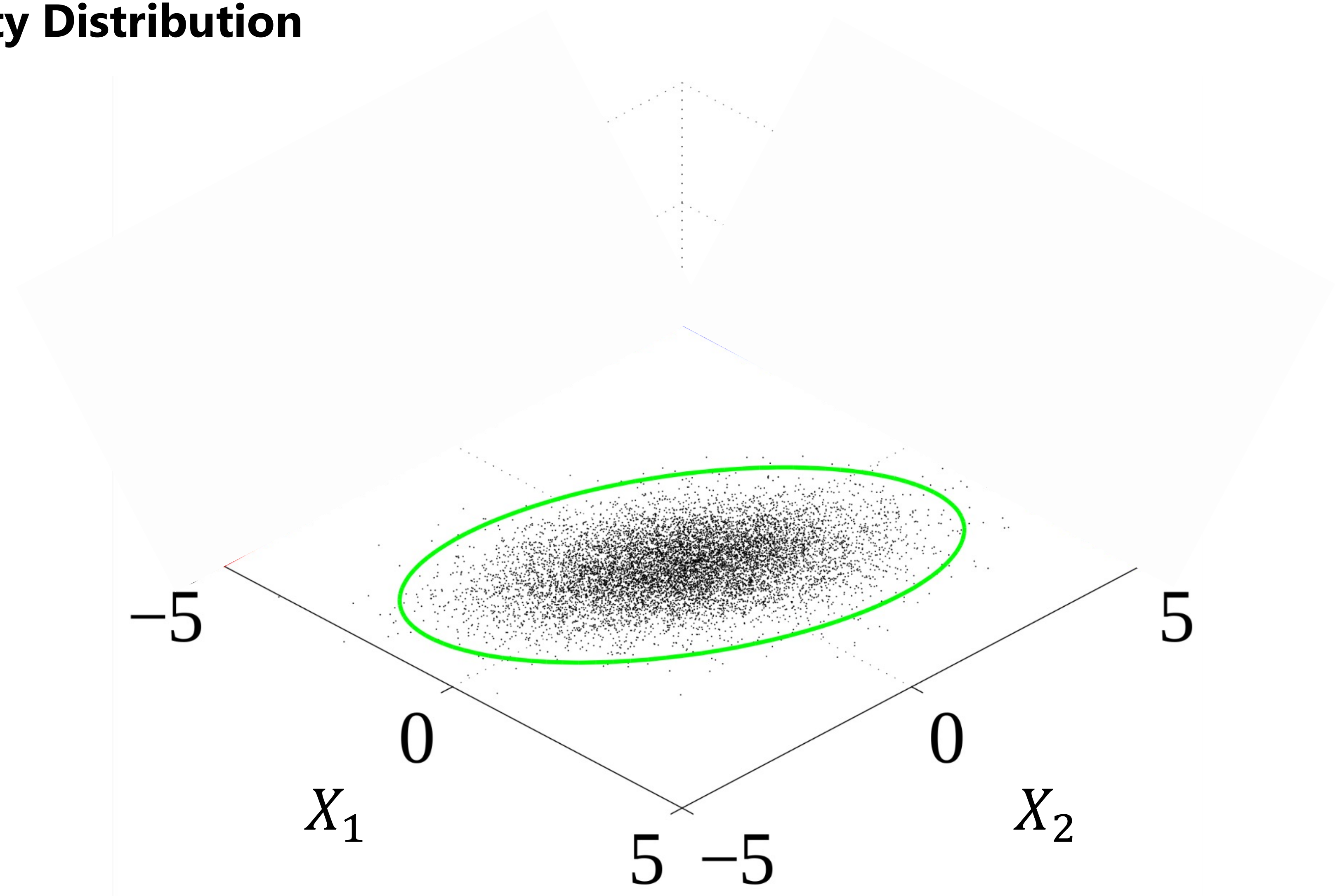


age	x_1
height	x_2
weight	x_3

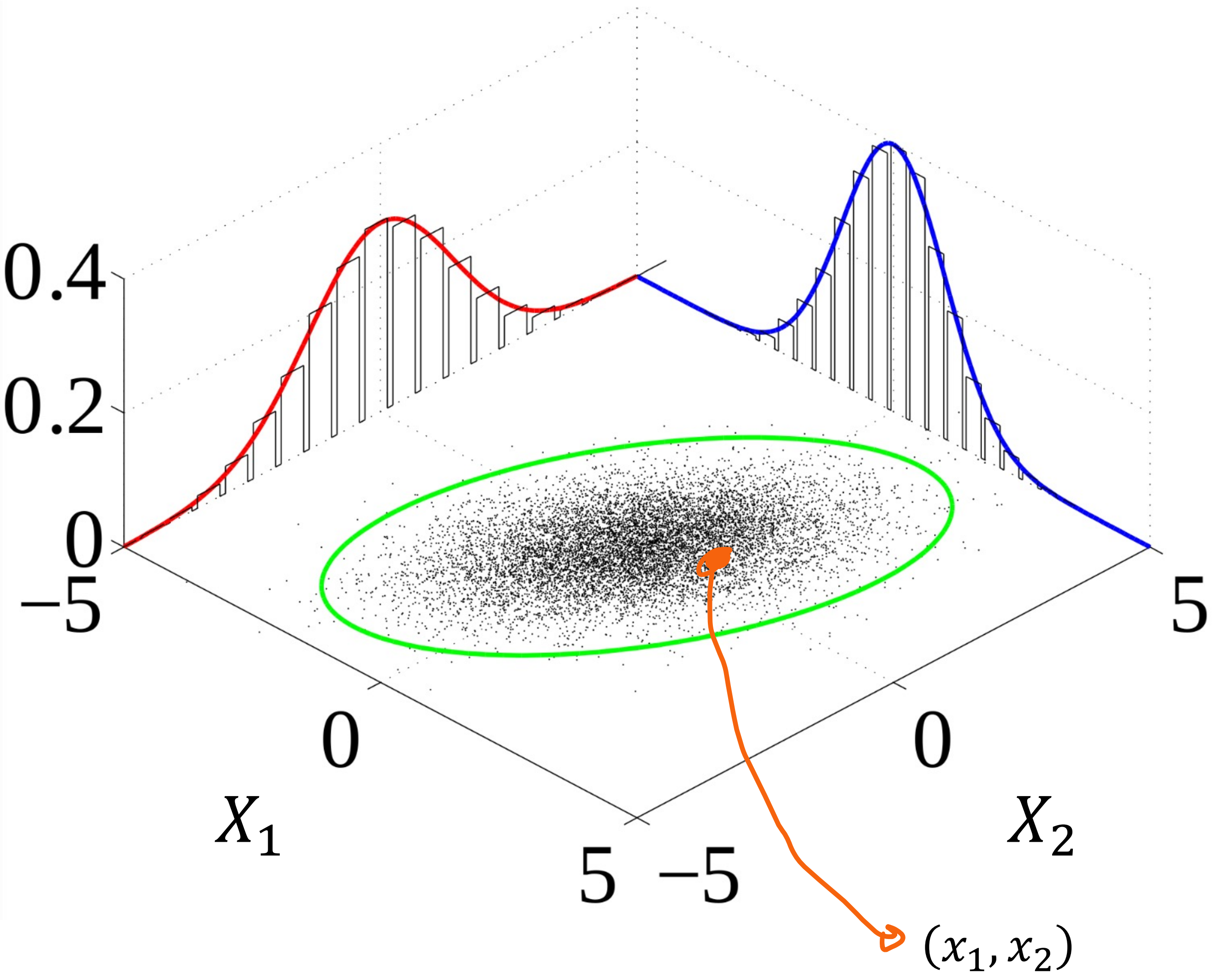
NB! *small* x

x denotes "*state*".

Joint Probability Distribution



Joint Probability Distribution



- Joint probability density function (PDF) of random variables X_1, X_2 is

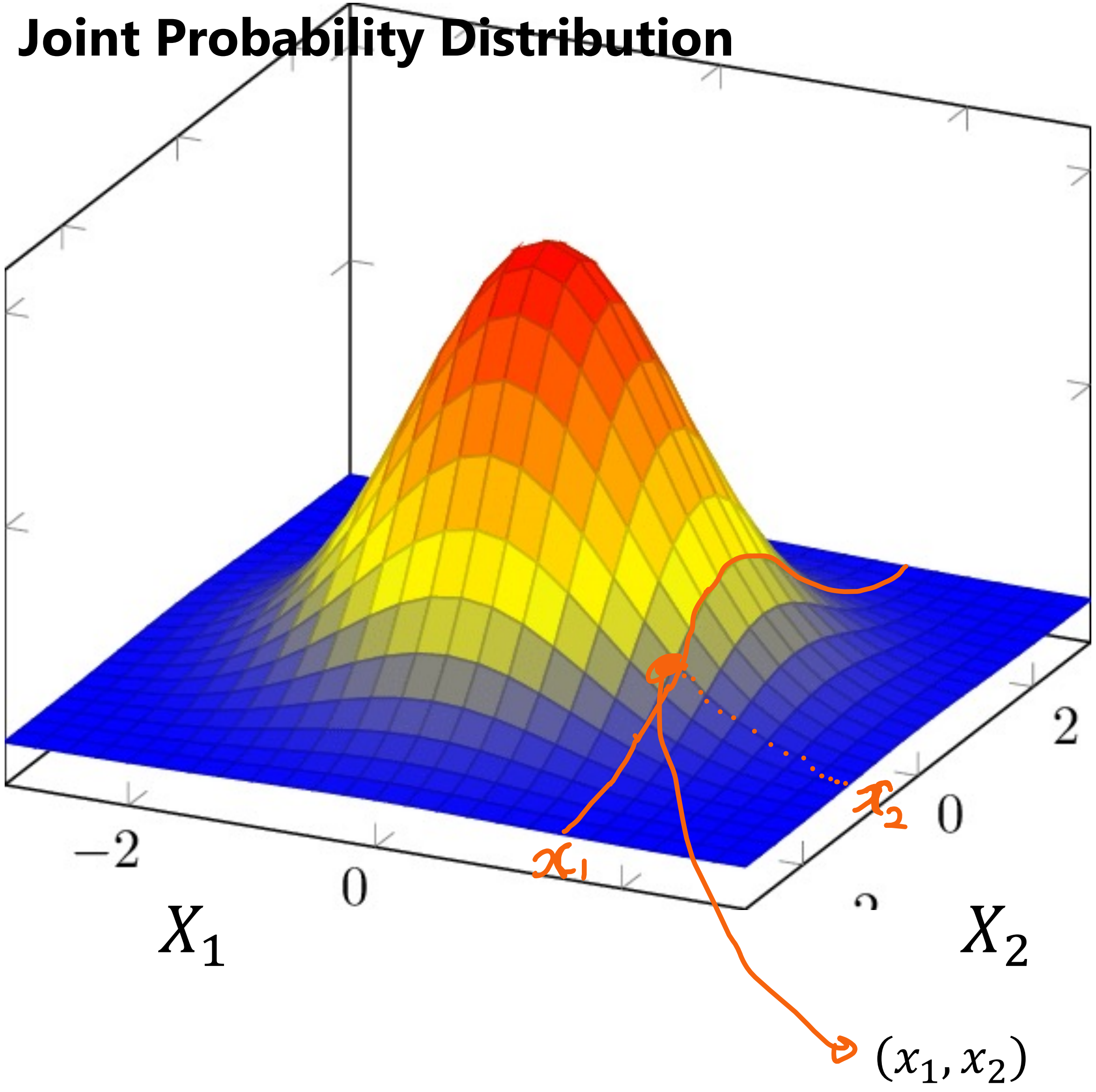
$$f_{X_1,X_2}(x_1, x_2)$$

- Marginal probability distribution:

$$f_{X_1}(x_1) = \int f_{X_1,X_2}(x_1, x_2)\, dx_2$$

$$f_{X_2}(x_2) = \int f_{X_1,X_2}(x_1, x_2)\, dx_1$$

Joint Probability Distribution



- Joint probability density function (PDF) of random variables X_1, X_2 is

$$f_{X_1,X_2}(x_1, x_2)$$

- Marginal probability distribution:

$$f_{X_1}(x_1) = \int f_{X_1,X_2}(x_1, x_2)\, dx_2$$

$$f_{X_2}(x_2) = \int f_{X_1,X_2}(x_1, x_2)\, dx_1$$

Random Vector

Moments of a distribution function

- In mathematics, the moments of a function are *certain quantitative measures related to the shape of the function's graph*.
- We will look at
  - E: Mean of random vector and random matrices.
  - Cov: Covariance matrix
  - Corr: Correlation matrix
  - E and Cov of multiple linear combinations.

Random Vector

Example: Boston Housing Price Dataset

- It is a database with information of areas around Boston city, and the median house prices.
- There are many variables, but we'll look at
  - {crime rate in the region, a number of rooms, home owner's age, housing price}

https://www.kaggle.com/code/andyxie/regression-with-r-boston-housing-price

Example: Boston Housing Price Dataset

- {crime rate in the region, a number of rooms, home owner's age, housing price}

````
```{r}
library(MASS)
library(ISLR)

summary(Boston)  # dataset description here:
https://www.kaggle.com/code/andyxie/regression-with-r-boston-housing-price

# convert the dataset to the DataFrame-type
Boston <- as.data.frame(Boston)

# select some features
df <- as.data.frame(Boston)[,c("crim","rm", "age", "medv")]

# plot
library(GGally)
ggpairs(df)
```
````

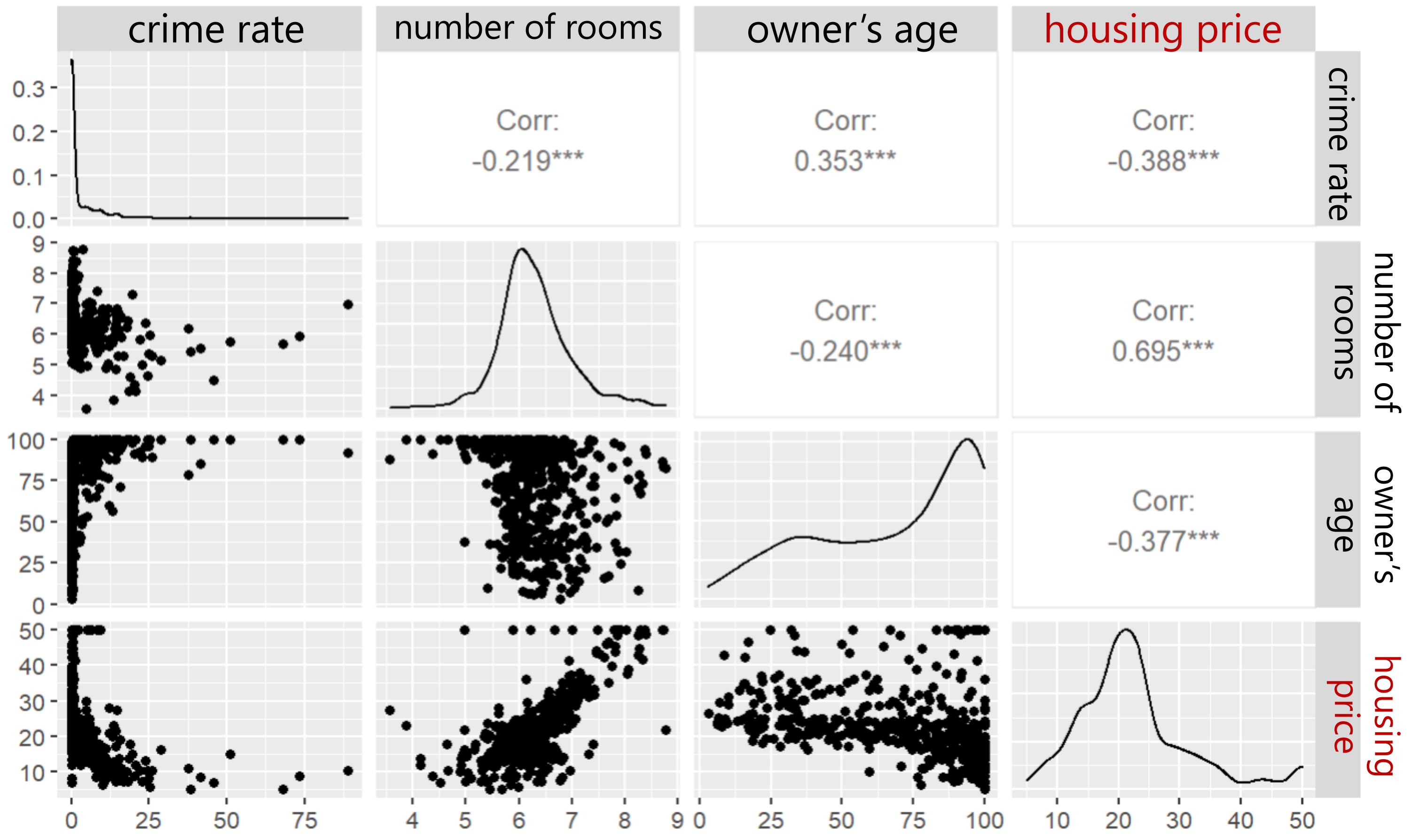
- **crim**: crime rate
- **rm**: number of rooms
- **age**: home owner's age
- **medv**: housing price

https://www.kaggle.com/code/andyxie/regression-with-r-boston-housing-price

Random Vector

Example: Boston Housing Price Dataset

- {crime rate in the region, a number of rooms, home owner's age, housing price}



Random Vector

Rules for means

- $\boldsymbol{X} = [X_1 \quad X_2 \quad X_3]$
- $\mathrm{E}(\boldsymbol{X}) = [\mathrm{E}(X_1) \quad \mathrm{E}(X_2) \quad \mathrm{E}(X_3)]$
- Given random matrix $\boldsymbol{X}_{(n \times p)}$ and random matrix $\boldsymbol{Y}_{(n \times p)}$

$$\mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$$

  - *Rules of vector/matrix addition*
  - n*: number of samples*
  - p*: number of covariates (input features)*

Random Vector

- Random matrix $\boldsymbol{X}_{(n\times p)}$ and conformable constant matrices (scalar matrices) $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\mathrm{E}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$$

- Matrix multiplication

$$\begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

- Proof:

$\boldsymbol{A} \in \mathbb{R}^{i\times k}, \boldsymbol{X} \in \mathbb{R}^{k\times l}, \boldsymbol{B} \in \mathbb{R}^{l\times j}$

$\boldsymbol{AXB} \in \mathbb{R}^{i\times j}$

e_{ij} is an element of $\boldsymbol{AXB}$

$e_{ij} = \sum_{k=1} a_{ik} \sum_{l=1} x_{kl} b_{lj}$

$\mathrm{E}[e_{ij}] = \sum_{k=1} a_{ik} \sum_{l=1} \mathrm{E}[x_{kl}] b_{lj}$

Random Vector

Q. What are the univariate analogue to the formula on the previous slide?

$\mathrm{E}[a\boldsymbol{X}b]$ instead of $\mathrm{E}[\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}]$ where a and b are constants.

$$\mathrm{E}[a\boldsymbol{X}b] = a\mathrm{E}[\boldsymbol{X}]b = ab\mathrm{E}[\boldsymbol{X}]$$

What about $\mathrm{E}[a\boldsymbol{X} + b\boldsymbol{Y}]$?

$$\mathrm{E}[a\boldsymbol{X} + b\boldsymbol{Y}] = a\mathrm{E}[\boldsymbol{X}] + b\mathrm{E}[\boldsymbol{Y}]$$

Covariance

- $\mathrm{Cov}(X_i, X_j) = \rho_{ij}$
  $= \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)]$
  $= \mathrm{E}[(X_i - \mu_i)] - \mu_i \mu_j$

$\mu = \mathrm{E}[X]$

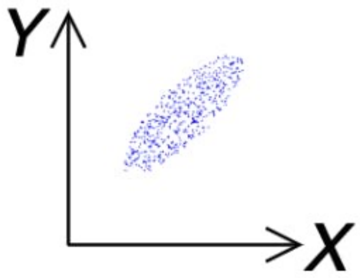
$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2\right] \\ &= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 \end{aligned}$$

- What's the covariance called when $i = j$?

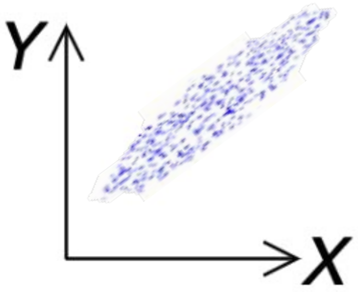
$$\mathrm{Cov}(X_i, X_i) = \mathrm{E}[(X_i - \mu_i)^2] = \mathrm{Var}(X_i)$$

NB! The cov values are not exact but approximate for the example purpose.

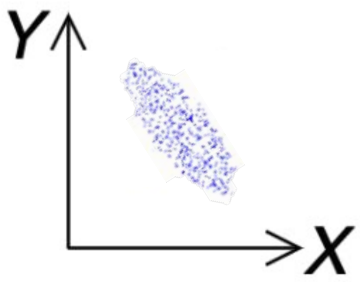
Example of $\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})$



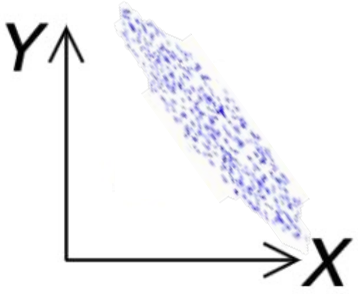
$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) = 1$



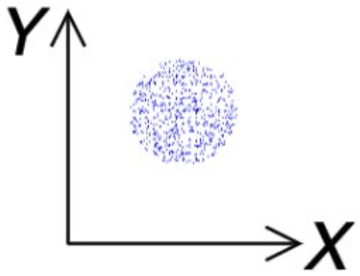
$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) = 2$



$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) = -1$



$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) = -2$



$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) = 0$

Variance-covariance matrix

- Variance-covariance matrix $\mathbf{\Sigma}$
- $\boldsymbol{X} \in \mathbb{R}^{p \times n}$
- $\mathbf{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] \in \mathbb{R}^{p \times p} = \mathrm{E}[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$

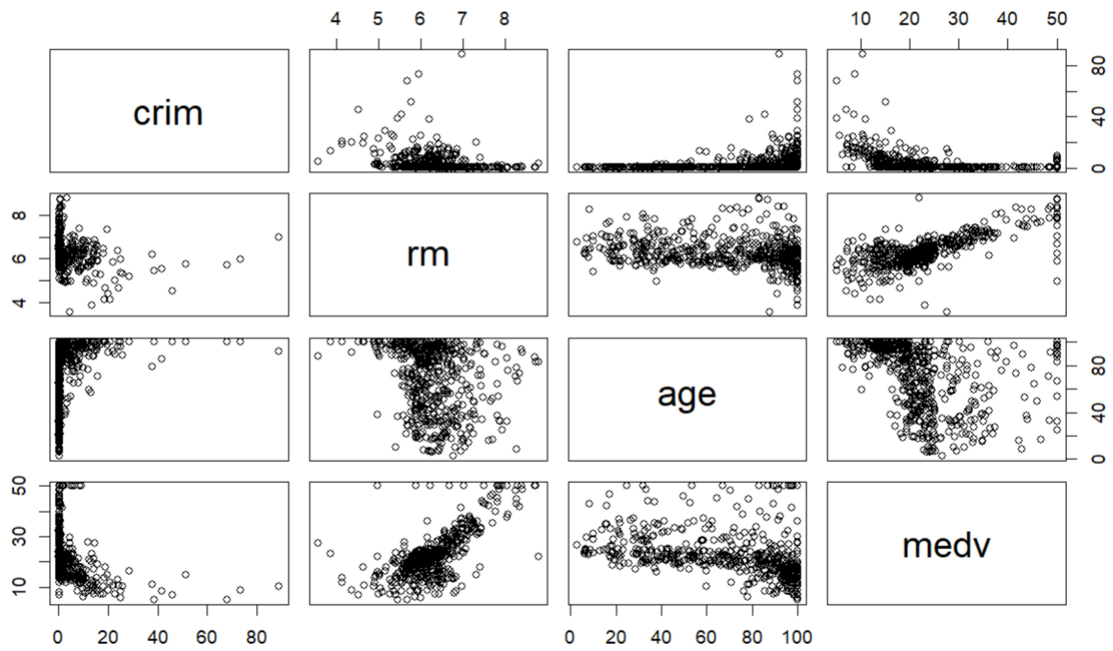
$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

covariance

variance

Variance-covariance matrix

crime rate | number of rooms | owner's age | housing price




```{r}
cov = cov(df)
```

```
           crim         rm         age        medv
crim  73.986578 -1.3250378  85.405322 -30.718508
rm    -1.325038  0.4936709  -4.751929   4.493446
age   85.405322 -4.7519292 792.358399 -97.589017
medv -30.718508  4.4934459 -97.589017  84.586724
```

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

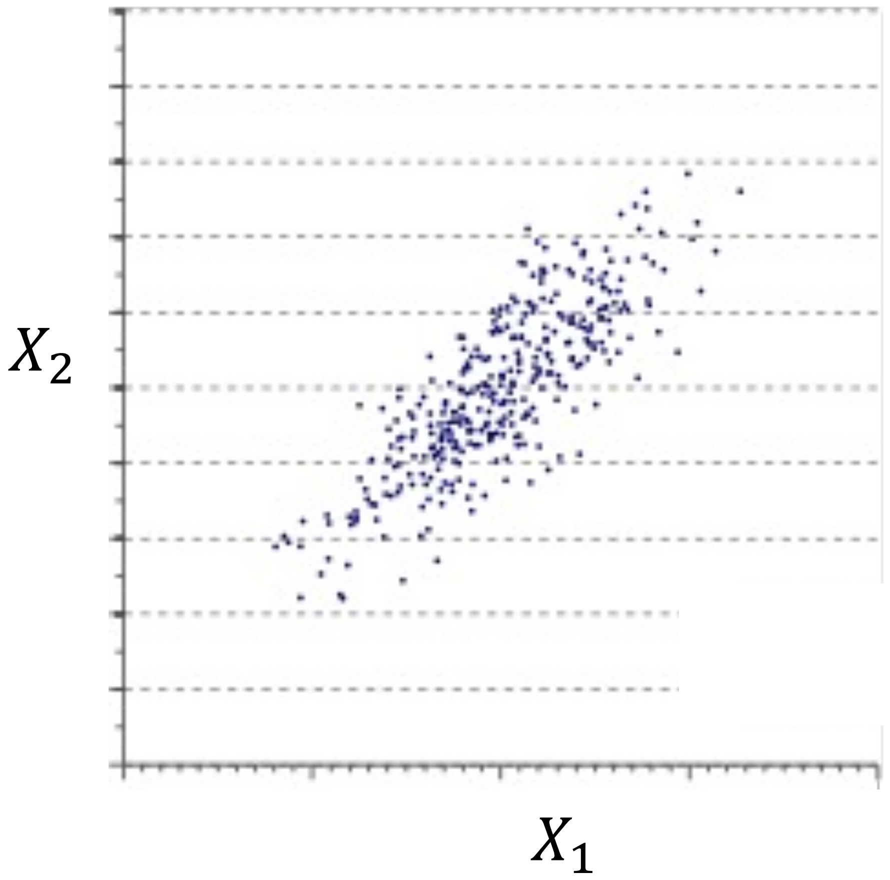
covariance

variance

Exercise: the variance-covariance matrix

- $\boldsymbol{X} \in \mathbb{R}^{p(=2)\times n} = [X_1, X_2]^T$
- $\mathrm{cov}(\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$



Q. What would be the correct Σ ?

① $\begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$

② $\begin{bmatrix} 5 & -4 \\ -4 & 6 \end{bmatrix}$

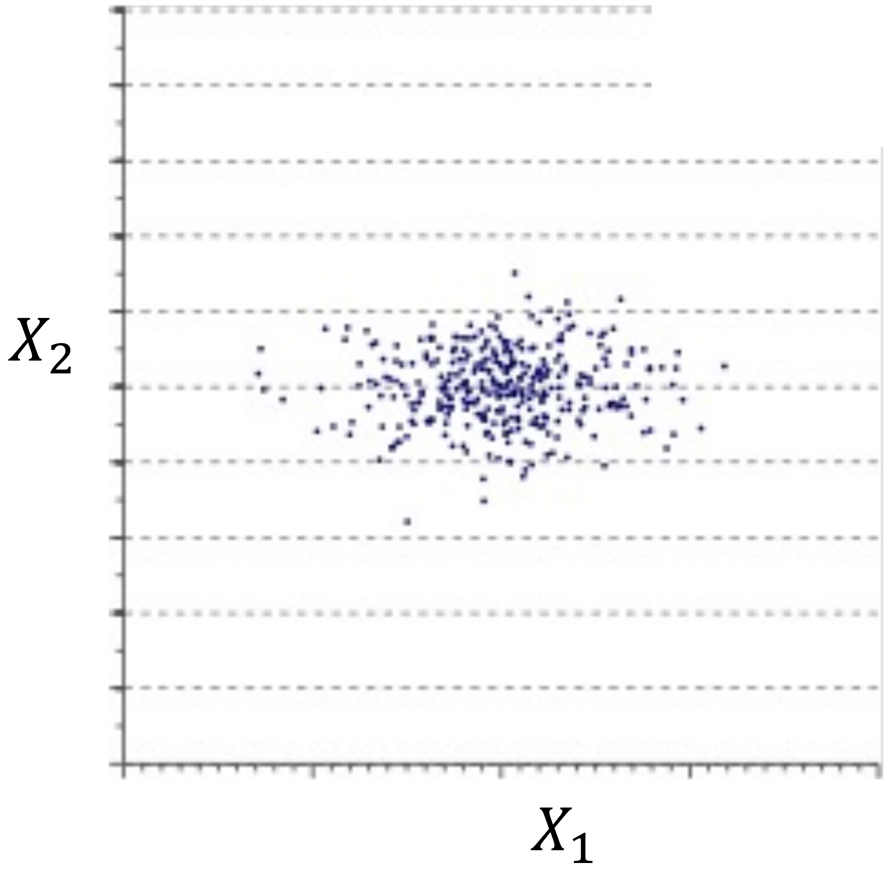
③ $\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$

④ $\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$

Exercise: the variance-covariance matrix

- $\boldsymbol{X} \in \mathbb{R}^{p(=2)\times n} = [X_1, X_2]^T$
- $\mathrm{cov}(\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$



Q. What would be the correct Σ ?

① $\begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$

② $\begin{bmatrix} 5 & -4 \\ -4 & 6 \end{bmatrix}$

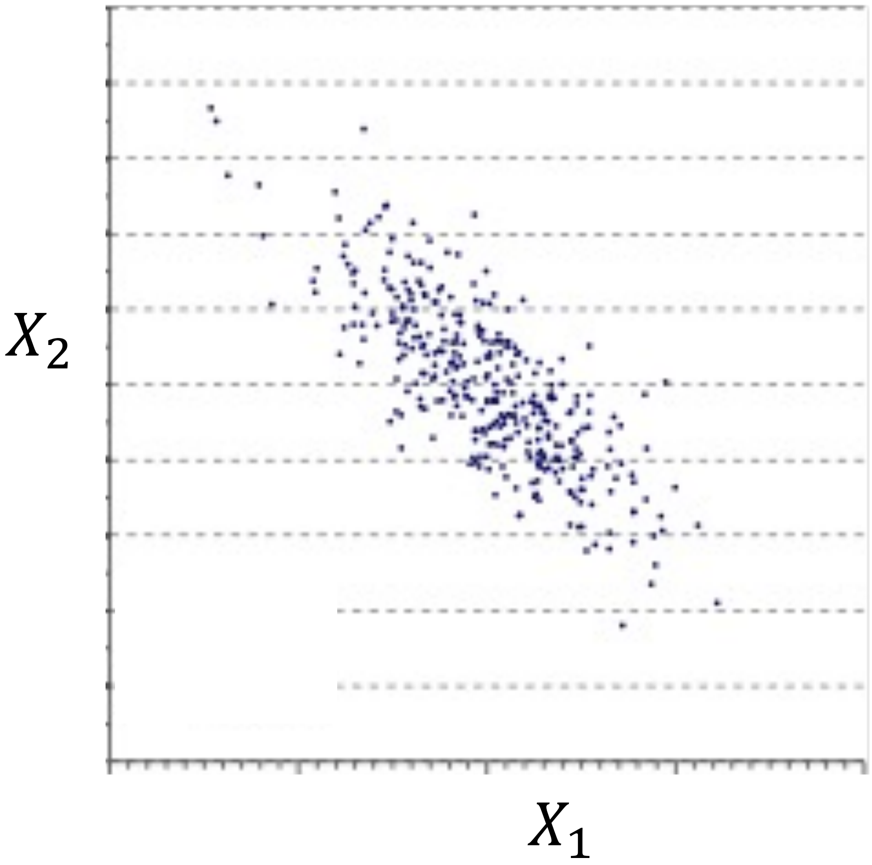
③ $\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$

④ $\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$

Exercise: the variance-covariance matrix

- $\boldsymbol{X} \in \mathbb{R}^{p(=2)\times n} = [X_1, X_2]^T$
- $\mathrm{cov}(\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$



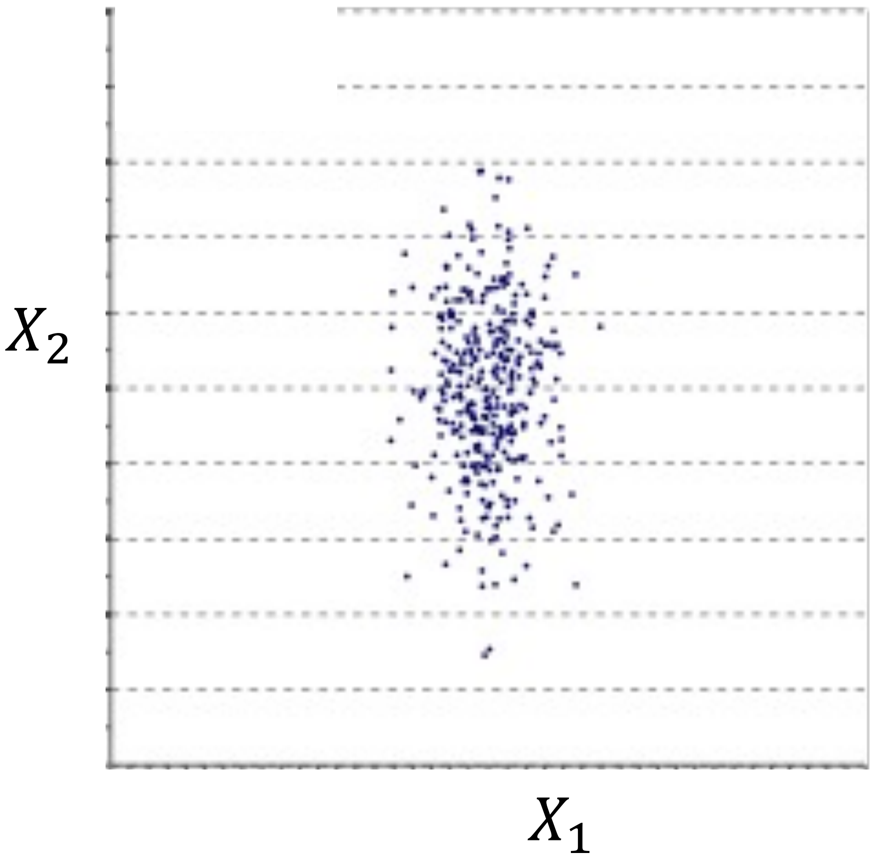
Q. What would be the correct Σ ?

1. $\begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$
2. $\begin{bmatrix} 5 & -4 \\ -4 & 6 \end{bmatrix}$
3. $\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$
4. $\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$

Exercise: the variance-covariance matrix

- $\boldsymbol{X} \in \mathbb{R}^{p(=2)\times n} = [X_1, X_2]^T$
- $\text{cov}(\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$



Q. What would be the correct Σ ?

1 $\begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$

2 $\begin{bmatrix} 5 & -4 \\ -4 & 6 \end{bmatrix}$

3 $\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$

4 $\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$

Random Vector

Correlation matrix

Correlation matrix $\boldsymbol{\rho}$ vs Covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_1^2}{\sqrt{\sigma_1^2\sigma_1^2}} & \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_1^2\sigma_p^2}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}} & \frac{\sigma_2^2}{\sqrt{\sigma_2^2\sigma_2^2}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_2^2\sigma_p^2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_1^2\sigma_p^2}} & \frac{\sigma_{2p}}{\sqrt{\sigma_2^2\sigma_p^2}} & \cdots & \frac{\sigma_p^2}{\sqrt{\sigma_p^2\sigma_p^2}} \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

- $\boldsymbol{\rho}$ is a normalized version of $\boldsymbol{\Sigma}$

Random Vector

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

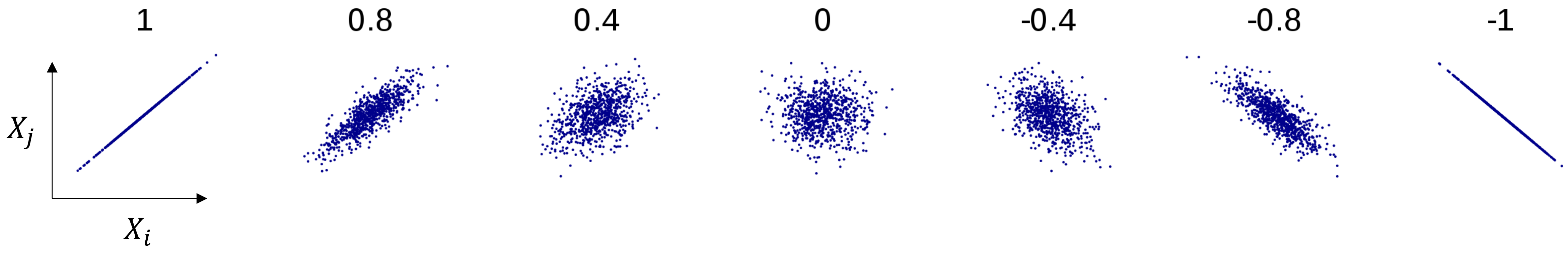
Correlation matrix

Correlation matrix $\boldsymbol{\rho}$ vs Covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_1^2}{\sqrt{\sigma_1^2\sigma_1^2}} & \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_1^2\sigma_p^2}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}} & \frac{\sigma_2^2}{\sqrt{\sigma_2^2\sigma_2^2}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_2^2\sigma_p^2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_1^2\sigma_p^2}} & \frac{\sigma_{2p}}{\sqrt{\sigma_2^2\sigma_p^2}} & \cdots & \frac{\sigma_p^2}{\sqrt{\sigma_p^2\sigma_p^2}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

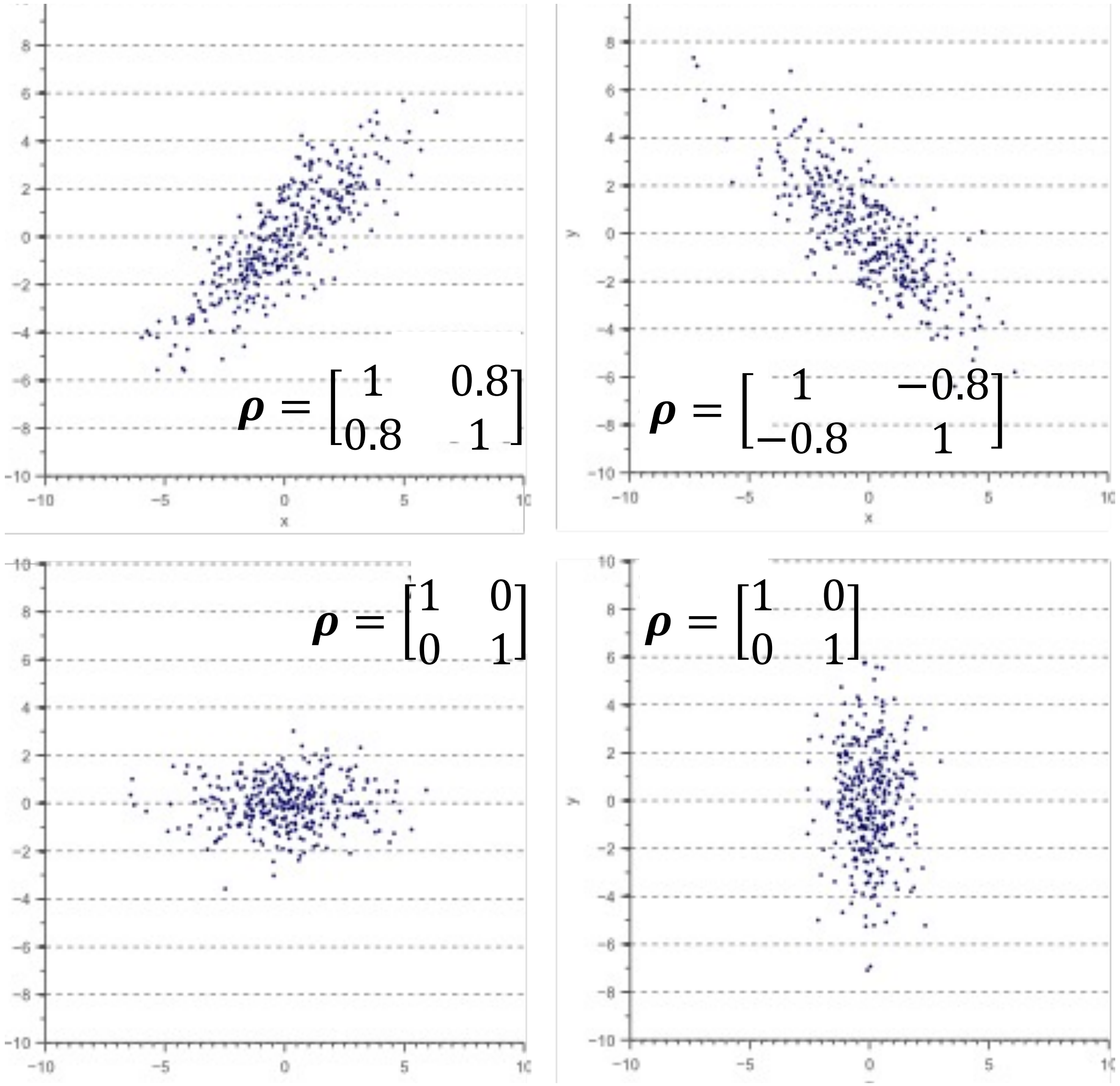
- $\boldsymbol{\rho}$ is a normalized version of $\boldsymbol{\Sigma}$
- $\rho \in [-1,1]$

How does ρ look?



Random Vector

How does ρ look?



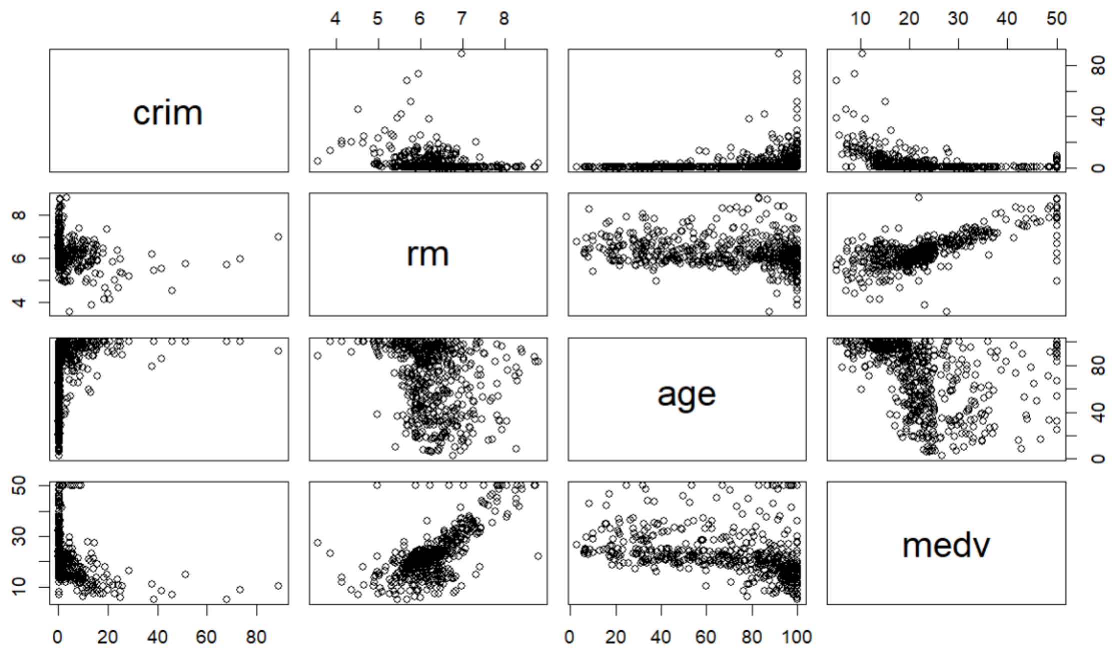
$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \qquad \boldsymbol{\rho} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\rho} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Random Vector

Correlation matrix on the Boston Housing price dataset

crime rate | number of rooms | owner's age | housing price




```{r}
cov = cov(df)
```

```
            crim         rm        age       medv
crim   73.986578 -1.3250378  85.405322 -30.718508
rm     -1.325038  0.4936709  -4.751929   4.493446
age    85.405322 -4.7519292 792.358399 -97.589017
medv  -30.718508  4.4934459 -97.589017  84.586724
```

```{r}
corr = cor(df)
corr
```

```
            crim         rm        age       medv
crim   1.0000000 -0.2192467  0.3527343 -0.3883046
rm    -0.2192467  1.0000000 -0.2402649  0.6953599
age    0.3527343 -0.2402649  1.0000000 -0.3769546
medv  -0.3883046  0.6953599 -0.3769546  1.0000000
```

Random Vector

Linear combinations

$\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{X}$

- scalar matrix $\boldsymbol{C} \in \mathbb{R}^{k \times p}$
- random matrix $\boldsymbol{X} \in \mathbb{R}^{p \times n}$

Then,

- $\mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\mathrm{E}(\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}}$
- $\mathrm{Cov}(\boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{C}^T$
    - $\boldsymbol{\mu}_{\boldsymbol{X}}$: $\boldsymbol{\mu}$ of $\boldsymbol{X}$
    - $\boldsymbol{\Sigma}_{\boldsymbol{X}}$: $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$

Proof of $\mathrm{Cov}(\boldsymbol{Z})$

$= \mathrm{E}[(\boldsymbol{Z} - \mathrm{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathrm{E}[\boldsymbol{Z}])^T]$

$= \mathrm{E}[(\boldsymbol{C}\boldsymbol{X} - \mathrm{E}[\boldsymbol{C}\boldsymbol{X}])(\boldsymbol{C}\boldsymbol{X} - \mathrm{E}[\boldsymbol{C}\boldsymbol{X}])^T]$

$= \mathrm{E}[(\boldsymbol{C}\boldsymbol{X} - \boldsymbol{C}\mathrm{E}[\boldsymbol{X}])(\boldsymbol{C}\boldsymbol{X} - \boldsymbol{C}\mathrm{E}[\boldsymbol{X}])^T]$

$= \mathrm{E}[(\boldsymbol{C}\boldsymbol{X} - \boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{C}\boldsymbol{X} - \boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}})^T]$

$= \mathrm{E}\left[\boldsymbol{C}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})\left(\boldsymbol{C}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})\right)^T\right]$

$= \mathrm{E}[\boldsymbol{C}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^T\boldsymbol{C}^T]$

$= \boldsymbol{C}\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^T]\boldsymbol{C}^T$

$= \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{C}^T$

Random Vector

Univariate case of $\mathrm{Cov}(\boldsymbol{Z})$

$\boldsymbol{Z} = c\boldsymbol{X}$

- scalar c
- random matrix $\boldsymbol{X} \in \mathbb{R}^{p\times n}$

Then,

- $\mathrm{Var}(c\boldsymbol{X}) = c^2\mathrm{Var}(\boldsymbol{X})$

Proof of $\mathrm{Var}(c\boldsymbol{X})$

$\mathrm{Var}(\boldsymbol{X}) = E[(\boldsymbol{X} - E[\boldsymbol{X}])^2]$

$\mathrm{Var}(c\boldsymbol{X}) = E[(c\boldsymbol{X} - E[c\boldsymbol{X}])^2]$

$= \mathrm{E}[(c\boldsymbol{X} - cE[\boldsymbol{X}])^2]$

$= \mathrm{E}\left[\left(c(\boldsymbol{X} - E[\boldsymbol{X}])\right)^2\right]$

$= \mathrm{E}[c^2(\boldsymbol{X} - E[\boldsymbol{X}])^2]$

$= c^2\mathrm{E}[(\boldsymbol{X} - E[\boldsymbol{X}])^2]$

$= c^2\mathrm{Var}(\boldsymbol{X})$

Exercise: Linear combinations

- $\boldsymbol{X} = [X_1, X_2, X_3, X_4]^T$
- Find $\boldsymbol{C}$ such that $\boldsymbol{CX} = \boldsymbol{Y}$
- $\boldsymbol{Y} = [X_1 - X_3 \quad X_2 + X_4 \quad (X_2 + X_4) - (X_1 + X_3)]^T$

$$\underbrace{\begin{bmatrix} X_1 - X_3 \\ X_2 + X_4 \\ (X_2 + X_4) - (X_1 + X_3) \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}}_{\boldsymbol{X}}$$

Exercise: Linear combinations

- $\boldsymbol{X} = [X_1, X_2, X_3, X_4]^T$
- Find $\boldsymbol{C}$ such that $\boldsymbol{CX} = \boldsymbol{Y}$
- $\boldsymbol{Y} = [X_1 - X_3 \quad X_2 + X_4 \quad (X_2 + X_4) - (X_1 + X_3)]^T$

$$\underbrace{\begin{bmatrix} X_1 - X_3 \\ X_2 + X_4 \\ (X_2 + X_4) - (X_1 + X_3) \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} 1 & 0 & -1 & 0 \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}}_{\boldsymbol{X}}$$

Random Vector

Exercise: Linear combinations

- $\boldsymbol{X} = [X_1, X_2, X_3, X_4]^T$
- Find $\boldsymbol{C}$ such that $\boldsymbol{CX} = \boldsymbol{Y}$
- $\boldsymbol{Y} = [X_1 - X_3 \quad X_2 + X_4 \quad (X_2 + X_4) - (X_1 + X_3)]^T$

$$\underbrace{\begin{bmatrix} X_1 - X_3 \\ X_2 + X_4 \\ (X_2 + X_4) - (X_1 + X_3) \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}}_{\boldsymbol{X}}$$

Then, $\mathrm{Cov}(\boldsymbol{Y})$?

$\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{C}\mathrm{Cov}(\boldsymbol{X})\boldsymbol{C}^T = \boldsymbol{C\Sigma_X C}^T$

- $\mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{CX}) = \boldsymbol{C}\mathrm{E}(\boldsymbol{X}) = \boldsymbol{C\mu_X}$
- $\mathrm{Cov}(\boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{CX}) = \boldsymbol{C\Sigma_X C}^T$
  - $\boldsymbol{\mu_X}$: $\boldsymbol{\mu}$ of $\boldsymbol{X}$
  - $\boldsymbol{\Sigma_X}$: $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$

Quiz

Quiz

Q1: Mean of sum

Q. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two bivariate random vectors with $\mathrm{E}(\boldsymbol{X}) = (1,2)^T$ and $\mathrm{E}(\boldsymbol{Y}) = (2,0)^T$.

What is $\mathrm{E}(\boldsymbol{X}+\boldsymbol{Y})$?

- A: $(1.5,\ 1)^T$
- B: $(3,\ 2)^T$
- C: $(-1,\ 2)^T$
- D: $(1,\ -2)^T$

Quiz

Q2: Mean of linear combination

Q. $\boldsymbol{X}$ is a 2-dimaensional random vector with $\mathrm{E}[\boldsymbol{X}] = (2, 5)^T$, and $\boldsymbol{b} = (0.5, 0.5)^T$ is a constant vector. What is $\mathrm{E}[\boldsymbol{b}^T\boldsymbol{X}]$?

- A: 3.5
- B: 7
- C: 2
- D: 5

Quiz

Q3: Covariance

Q. $\boldsymbol{X}$ is a p-dimensional random vector with mean $\boldsymbol{\mu}$, and $\boldsymbol{X} \in \mathbb{R}^{p \times n}$. Which of the following defines the covariance matrix?

- A: $\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})^T (\boldsymbol{X} - \boldsymbol{\mu})]$
- B: $\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T]$
- C: $\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})]$
- D: $\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})^T (\boldsymbol{X} - \boldsymbol{\mu})^T]$

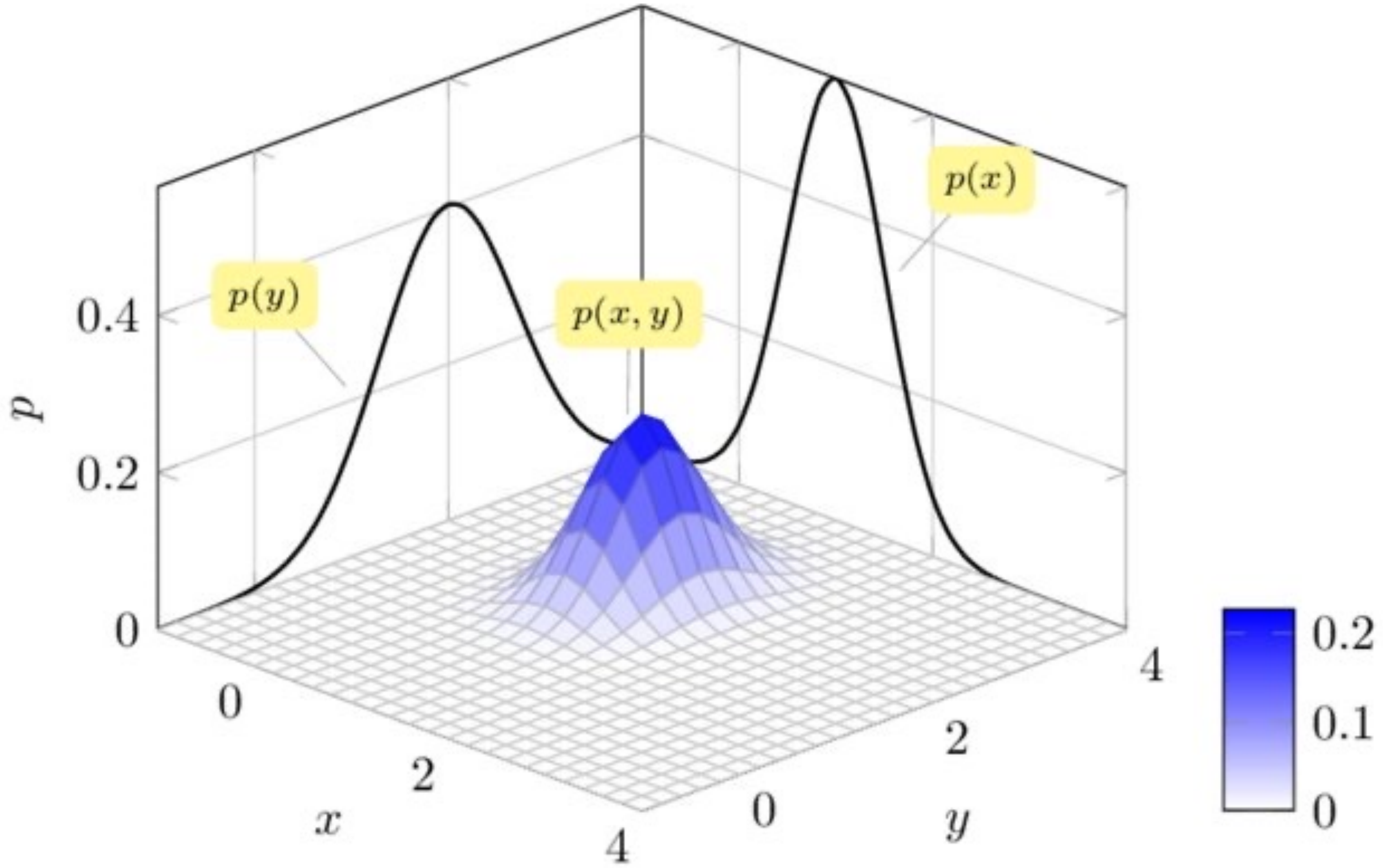
Multivarite Normal Distribution

Multivarite Normal (mvN) Distribution

Why is the mvN distribution so popular?

- Many natural phenomena can be modeled with it.
- Good interpretability of covariance
- Mathematically tractable
- Used as a building block in many methods

Multivarite Normal (mvN) Distribution



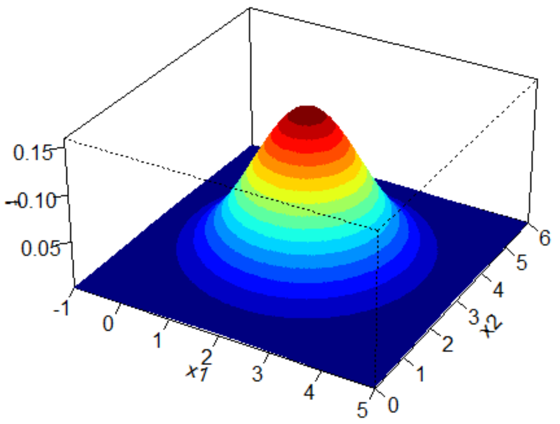
Multivarite Normal Distribution

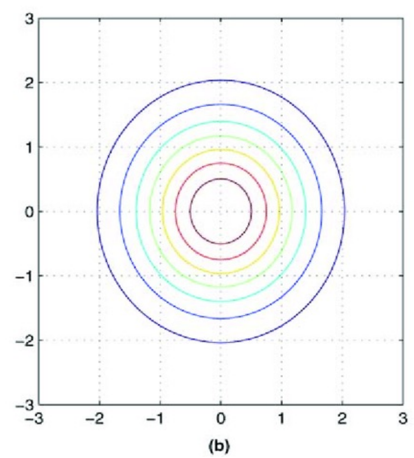
Multivariate Normal Probability Density Function (PDF)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

$$\boxed{f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}}$$

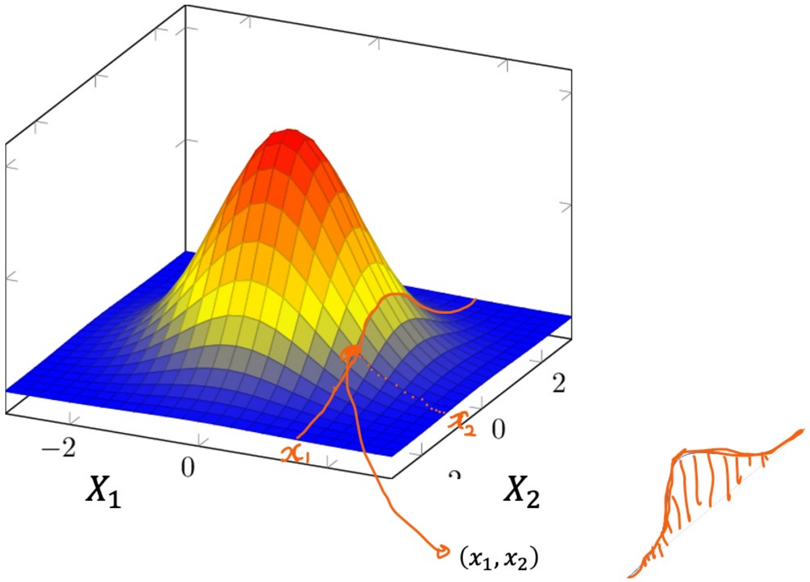
Multivarite Normal Distribution





Useful properties of the mvN distribution

- The graphical contours of the mvN are ellipsoids:
- Linear combinations of components of $\boldsymbol{X}$ are (multivariate) normal.
  - *e.g.*, If X_1 and X_2 are normal and indepdent, then $X_1 + X_2$ is also normal,.
- All subsets of components of $\boldsymbol{X}$ are (multivariate) normal.
- Zero covariance implies that the correponding components are independently distributed.



Multivarite Normal Distribution

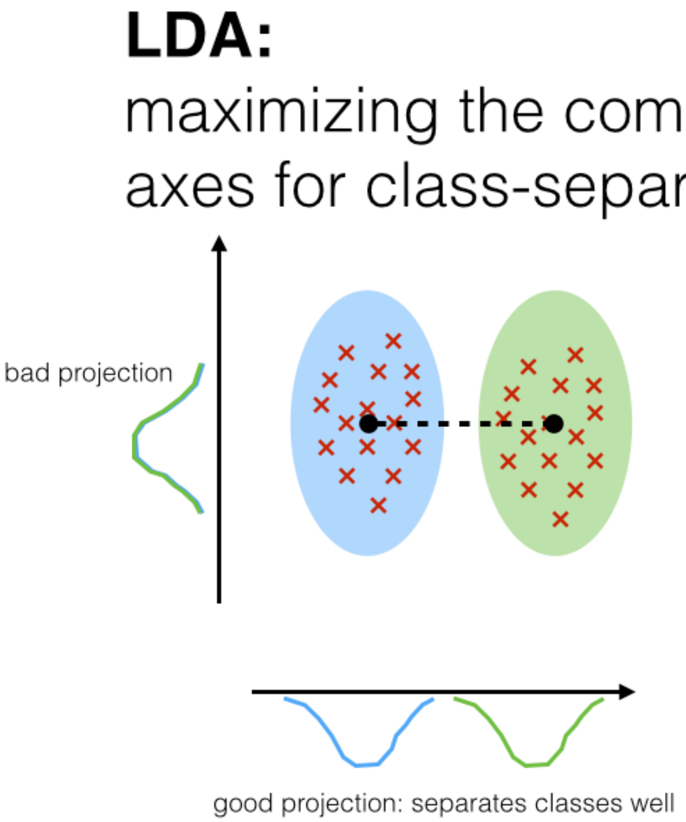
Some note

- All of these are proven in TMA2467 Linear Statistical Models.

Multivarite Normal Distribution

Wait! Why are we learning it?

- In M4 (classification), the mvN is very important!



LDA is a *supervised* linear transformation technique, to maximize the separation between different classes.

Quiz

Quiz

Question 1: Multivariate Normal PDF

The mvN distribution is expressed as

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2} \boxed{\qquad\qquad} \}$$

Q. What is $\boxed{\qquad\qquad}$?

- A: $(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$
- B: $(\boldsymbol{x} - \boldsymbol{\mu}) \boldsymbol{\Sigma} (\boldsymbol{x} - \boldsymbol{\mu})^T$
- C: $\boldsymbol{\Sigma} - \boldsymbol{\mu}$

Quiz

Question 2: Multivariate normal distribution

$\boldsymbol{X}_p \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{C}$ is a $k \times p$ constant matrix. $\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X}$ is

- A: Chi-squared with k degrees of freedom
- B: Multivariate normal with mean $k\boldsymbol{\mu}$
- C: Chi-squared with p degrees of freedom
- D: Multivariate normal with mean $\boldsymbol{C}\boldsymbol{\mu}$

Question 3: Independence

Let $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & 2 & 5 \end{bmatrix}.$$

Which two variables are independent?

- A: X_1 and X_2
- B: X_1 and X_3
- C: X_2 and X_3
- D: None – but two are uncorrelated.

Question 4: Constructing independent variables?

Let $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. How can I construct a vector of independent standard normal variables from $\boldsymbol{X}$?

- A: $\boldsymbol{\Sigma}(\boldsymbol{X} - \boldsymbol{\mu})$
- B: $\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} + \boldsymbol{\mu})$
- C: $\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$
- D: $\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{X} + \boldsymbol{\mu})$


The End!