

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

## TMA4268 Statistical Learning V2023

Daesoo Lee,

Department of Mathematical Sciences, NTNU



Norwegian University of  
Science and Technology

19/01/2023

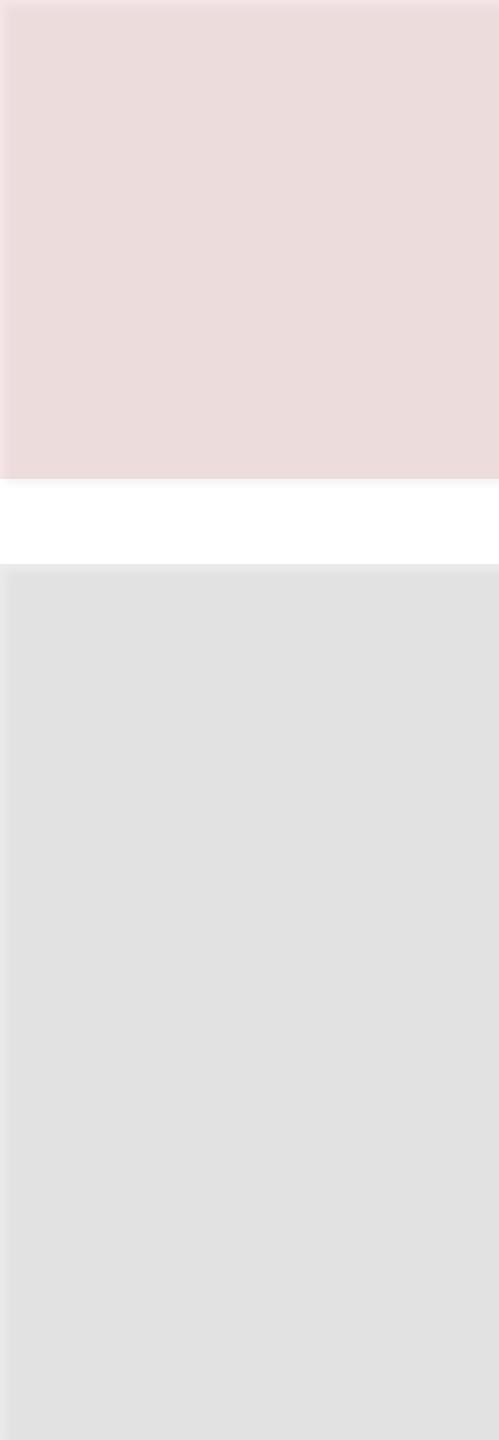
DA

## 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)

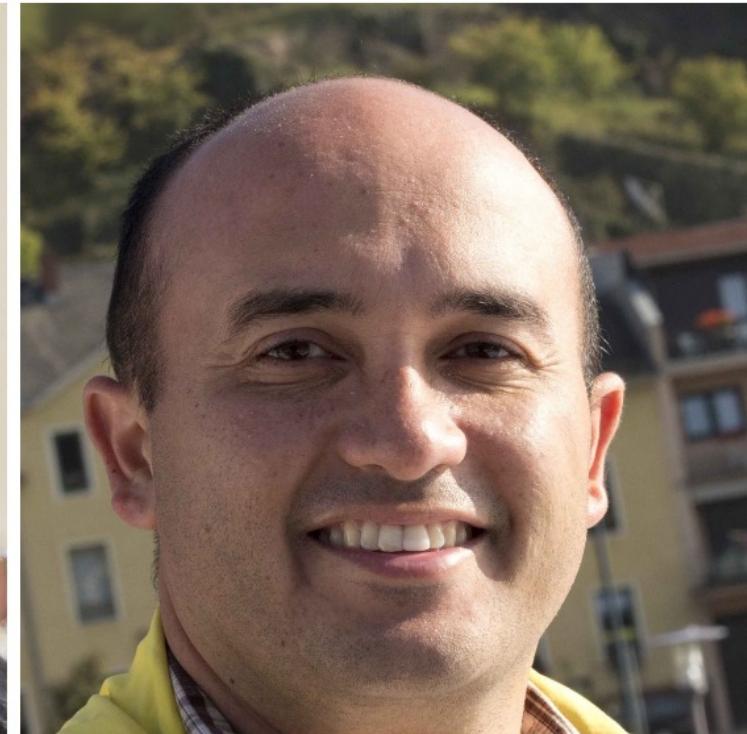
# Quiz

<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(?)

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

MSE: mean squared error

# Course Schedule

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

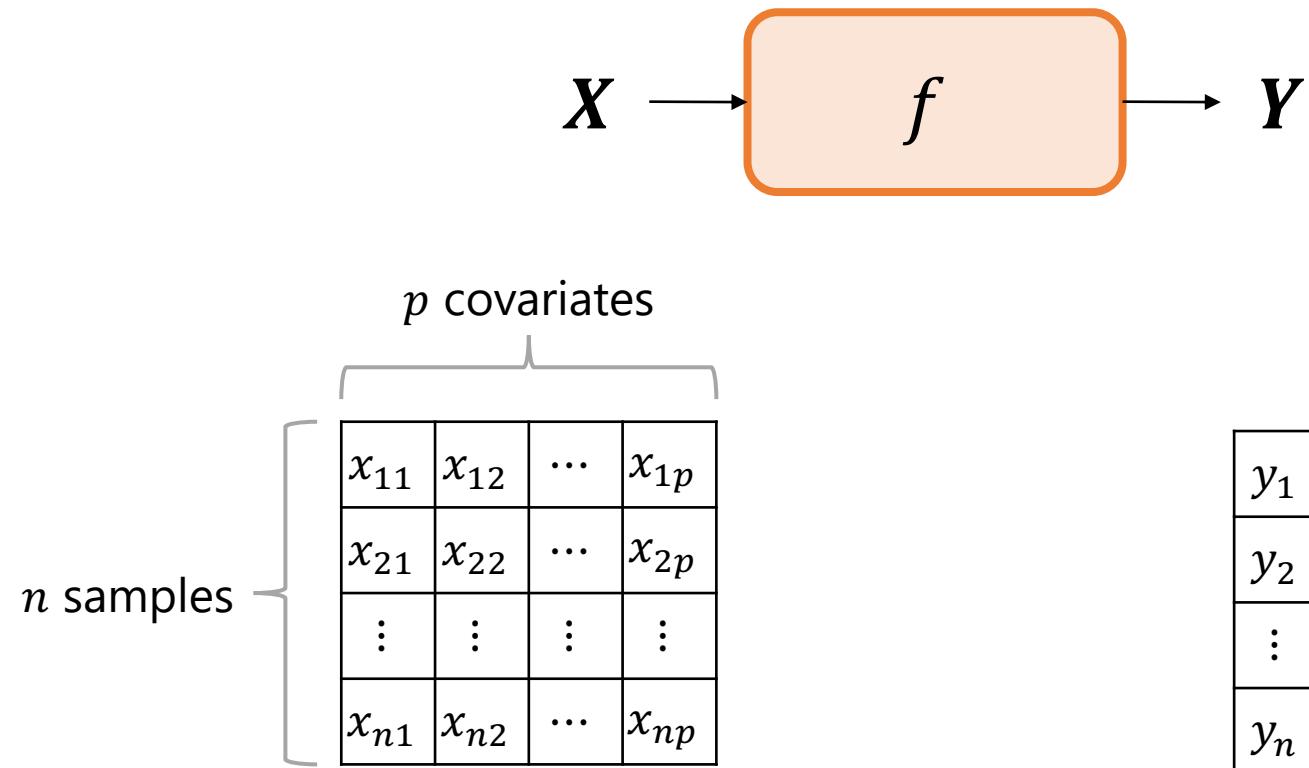
# Overview

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

# Random Vector

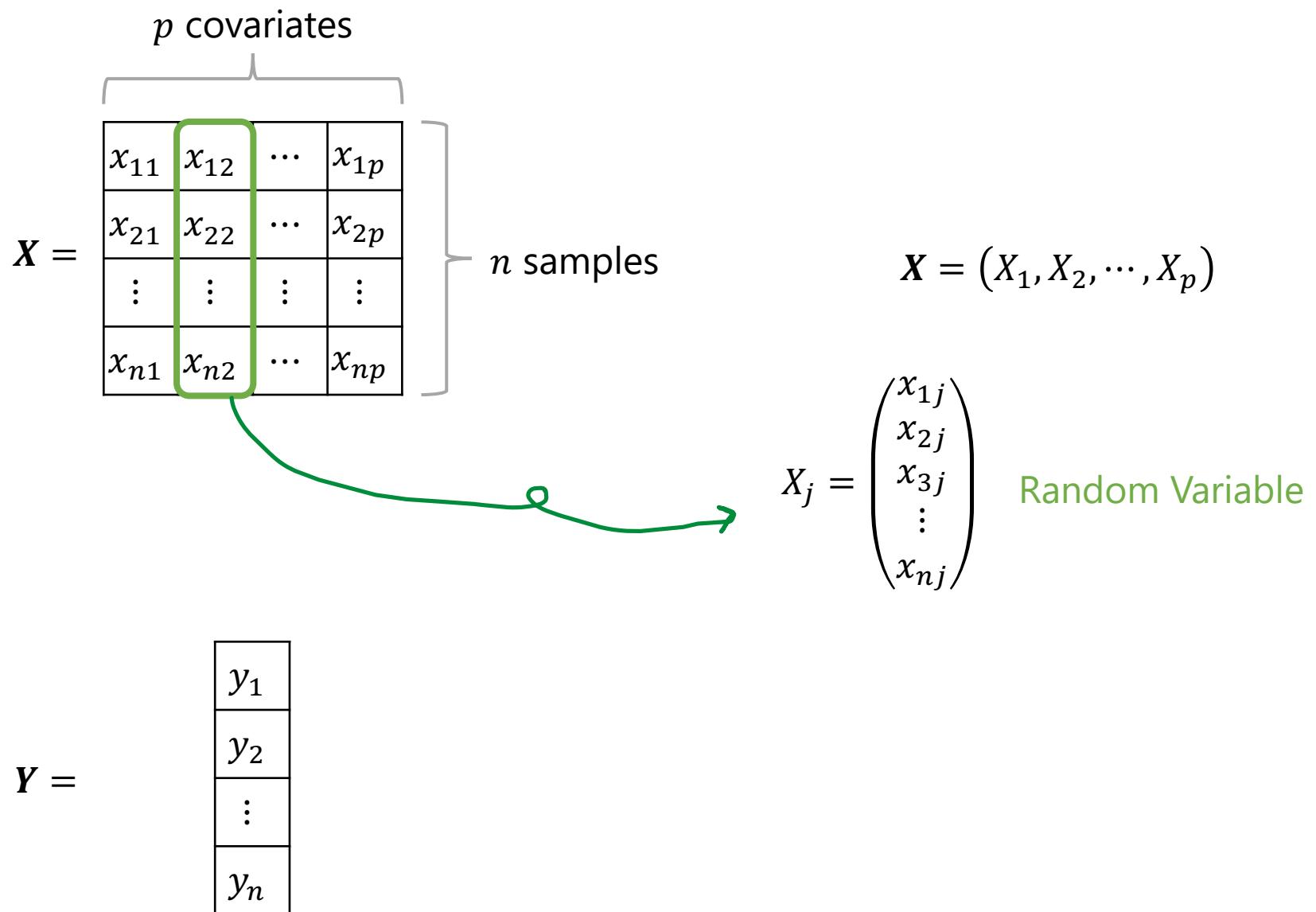
# Random Vector

## Recap



# Random Vector

## Recap



# Random Vector

## Random vector

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

age	$X_1$
height	$X_2$
weight	$X_3$



# Random Vector

## Random vector

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

age	$x_1$
height	$x_2$
weight	$x_3$

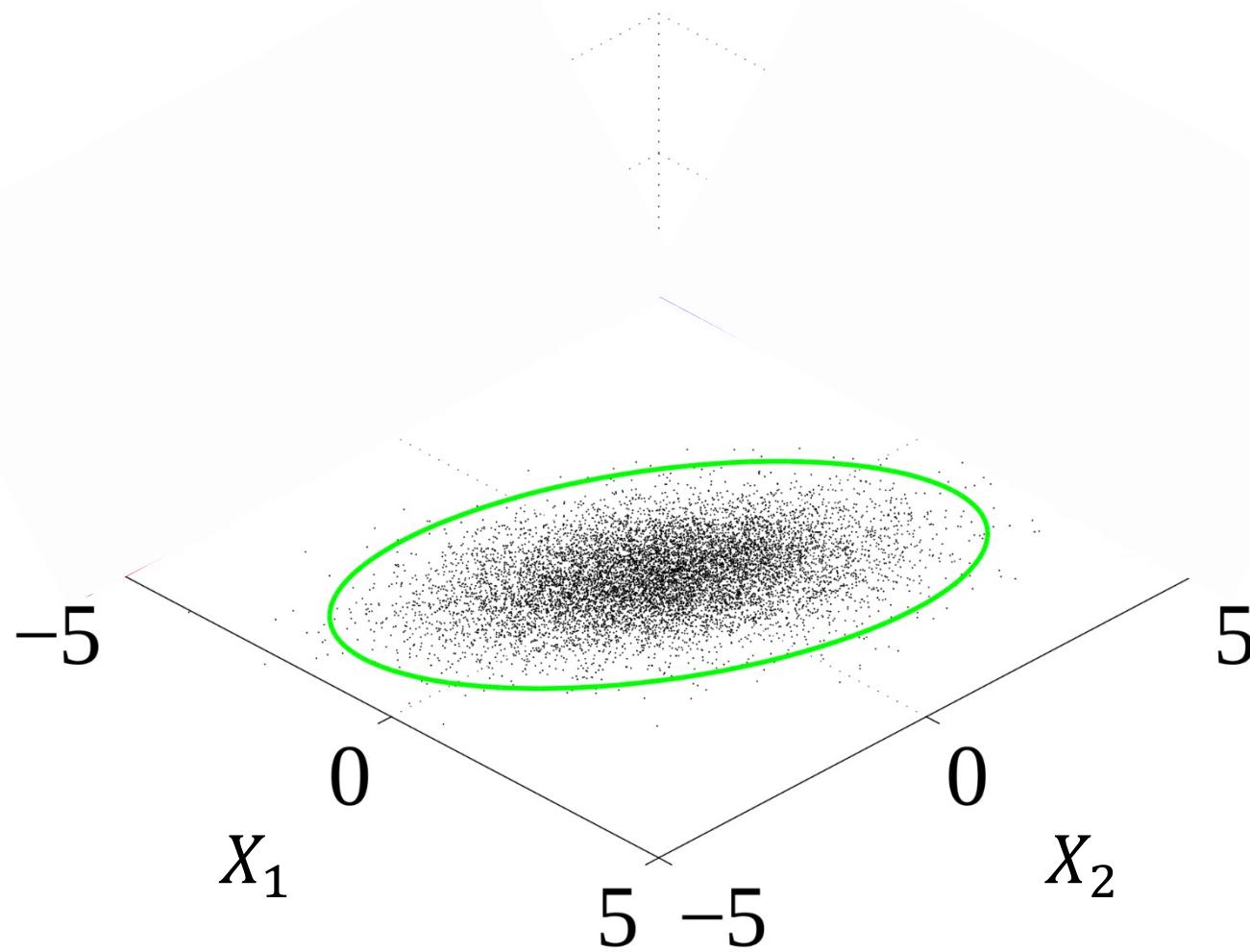
NB! *small x*

$x$  denotes "state".



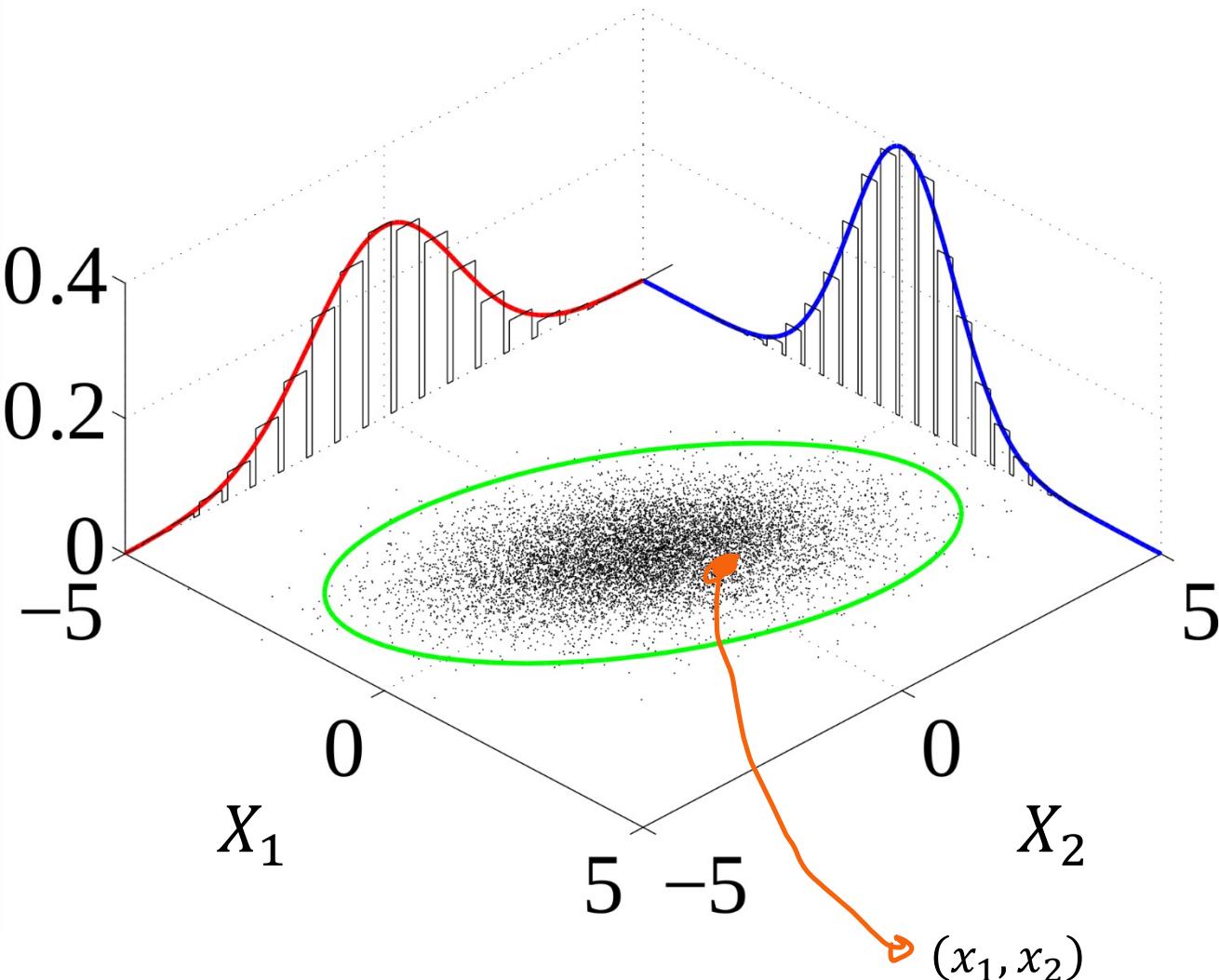
# Random Vector

## Joint Probability Distribution



# Random Vector

## Joint Probability Distribution



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

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

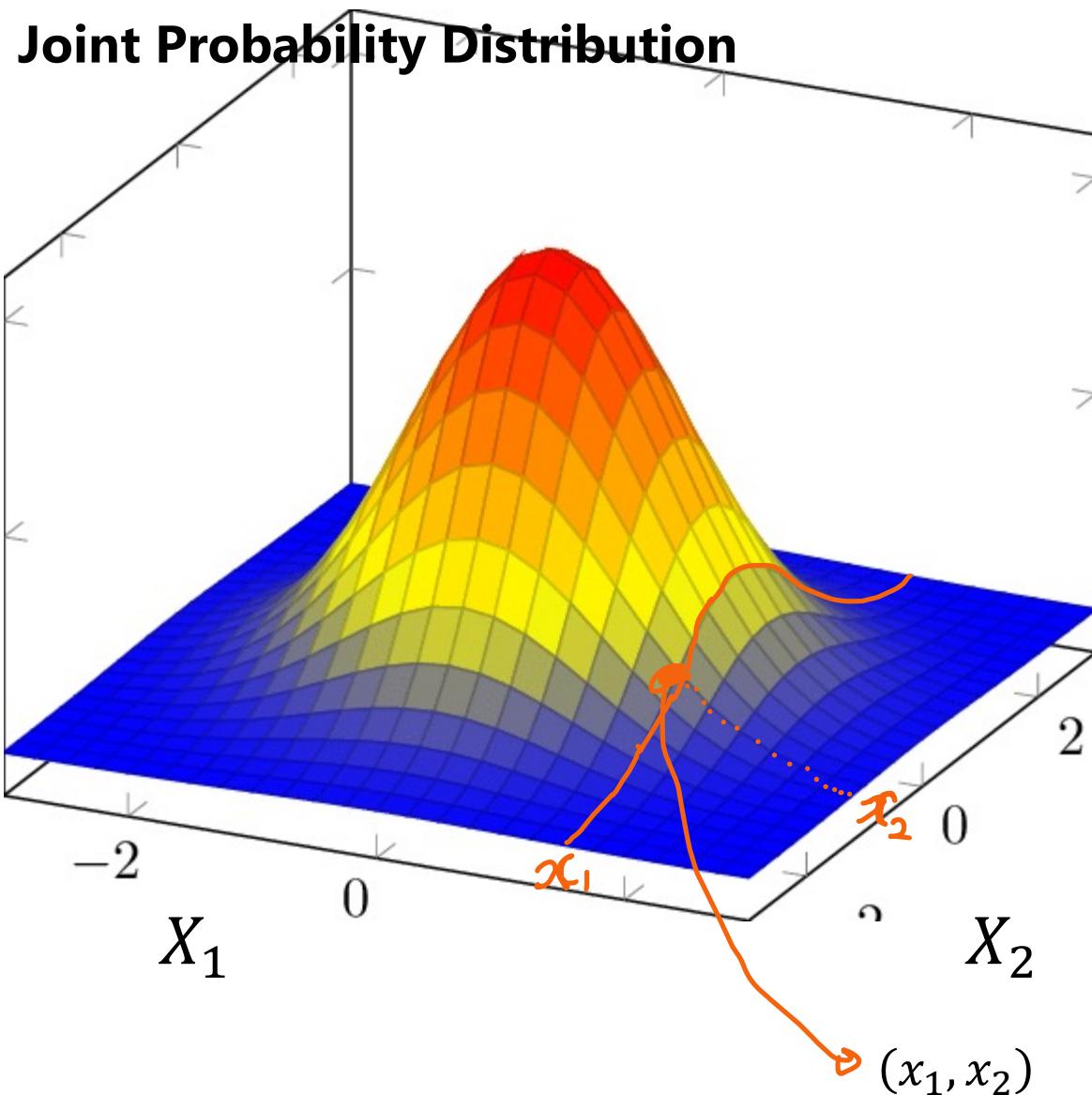
- Marginal probability distribution:

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

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

# Random Vector

## Joint Probability Distribution



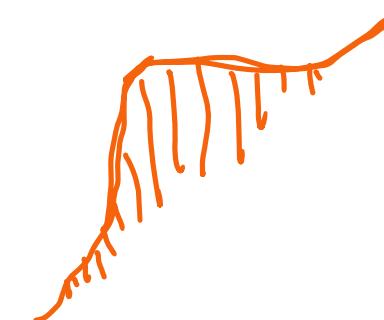
- Joint probability density function (PDF) of random variables  $X_1, X_2$  is

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

- Marginal probability distribution:

$$\underbrace{f_{X_1}(x_1)}_{\text{Marginal PDF}} = \int f_{X_1, X_2}(x_1, x_2) dx_2$$

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



## 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. (Expectation)
  - Cov: Covariance matrix
  - Corr: Correlation matrix
  - E and Cov of multiple linear combinations.

## Example: Boston Housing Price Dataset

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

# Random Vector

## 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)

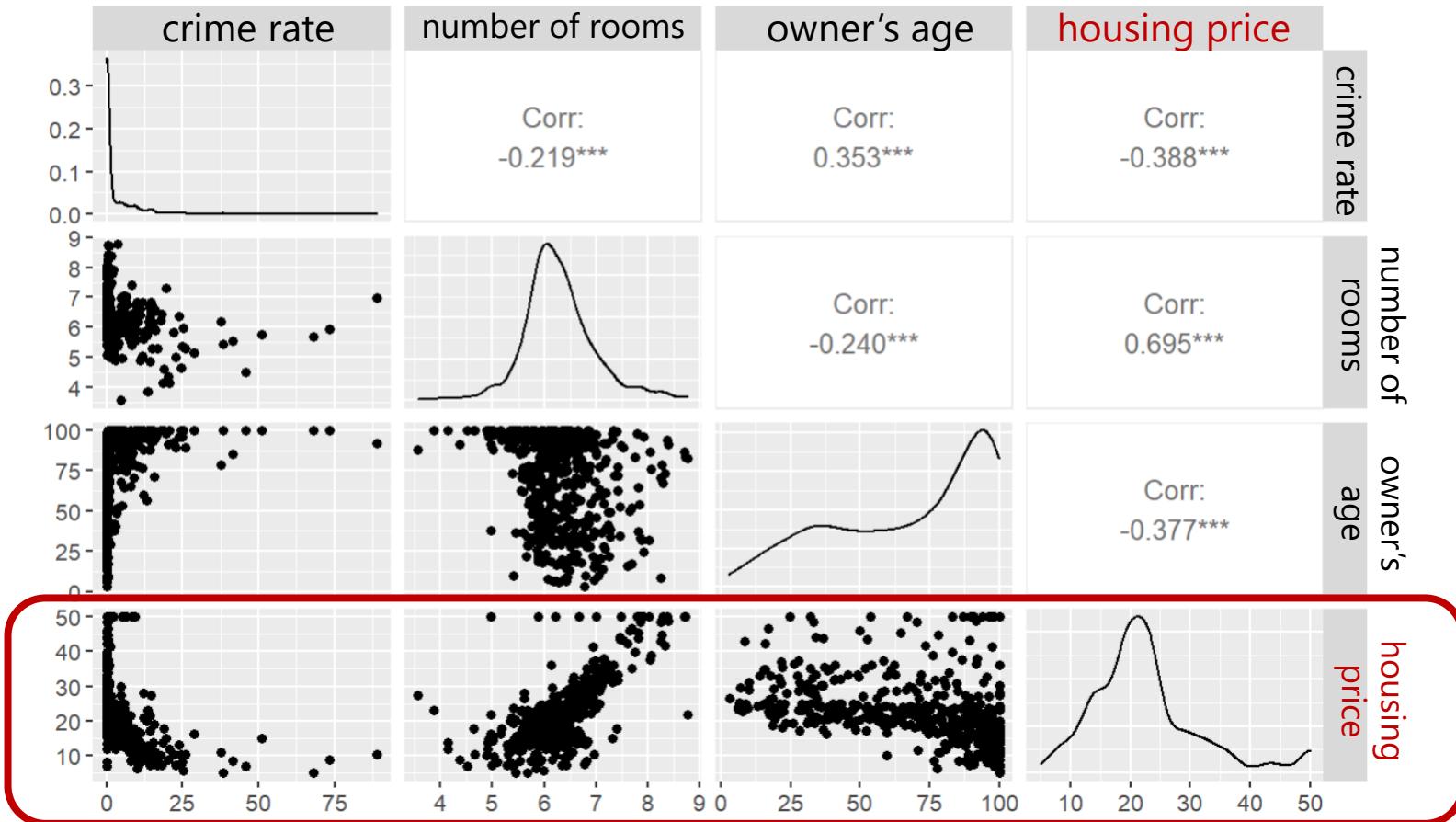
# select some features
df <- as.data.frame(Boston)[,c("crim", "rm", "age", "medv")]

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

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

# Random Vector

## Example: Boston Housing Price Dataset



## Rules for means

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

$$\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$$

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

# Random Vector

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

$$\mathbb{E}(\mathbf{AXB}) = \mathbf{A}\mathbb{E}(\mathbf{X})\mathbf{B}$$

- 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:

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

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

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

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

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

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

$E[aXb]$  instead of  $E[AXB]$  where  $a$  and  $b$  are constants.

$$E[aXb] = aE[X]b = abE[X]$$

What about  $E[aX + bY]$ ?

$$E[aX + bY] = aE[X] + bE[Y]$$

## Covariance

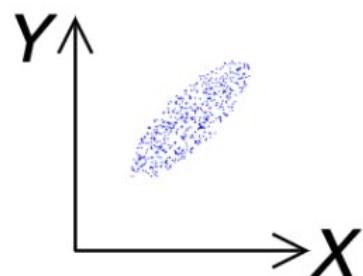
- $\text{Cov}(X_i, X_j) = \rho_{ij}$   
=  $E[(X_i - E[X_i])(X_j - E[X_j])]$   
=  $E[X_i X_j - X_i E[X_j] - E[X_i] X_j + E[X_i] E[X_j]]$   
=  $E[X_i X_j] - E[X_i] E[X_j] - E[X_i] E[X_j] + E[X_i] E[X_j]$   
=  $E[X_i X_j] - E[X_i] E[X_j]$   
=  $E[(X_i - \mu_i)] - \mu_i \mu_j$        $\mu = E[X]$

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

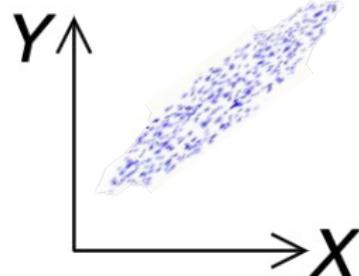
$$\text{Cov}(X_i, X_i) = E[(X_i - \mu_i)^2] = \text{Var}(X_i)$$

# Random Vector

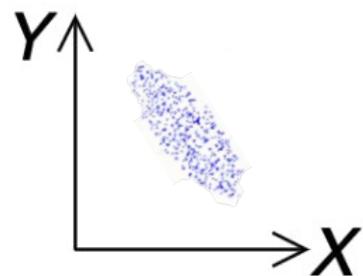
## Example of $\text{Cov}(X, Y)$



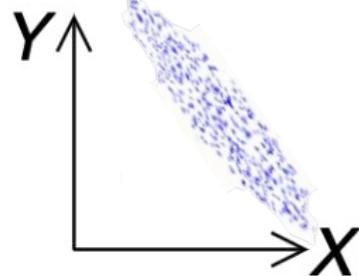
$$\text{Cov}(X, Y) = 1$$



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

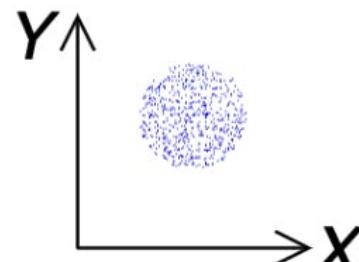


$$\text{Cov}(X, Y) = -1$$



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

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



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

# Random Vector

## Variance-covariance matrix

- Variance-covariance matrix  $\Sigma$
- $X \in \mathbb{R}^{p \times n}$ ; random matrix (instead of random vector)
- $\Sigma = \text{Cov}(X) = E[(X - \mu)(X - \mu)^T] \in \mathbb{R}^{p \times p} = E[XX^T] - \mu\mu^T$

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

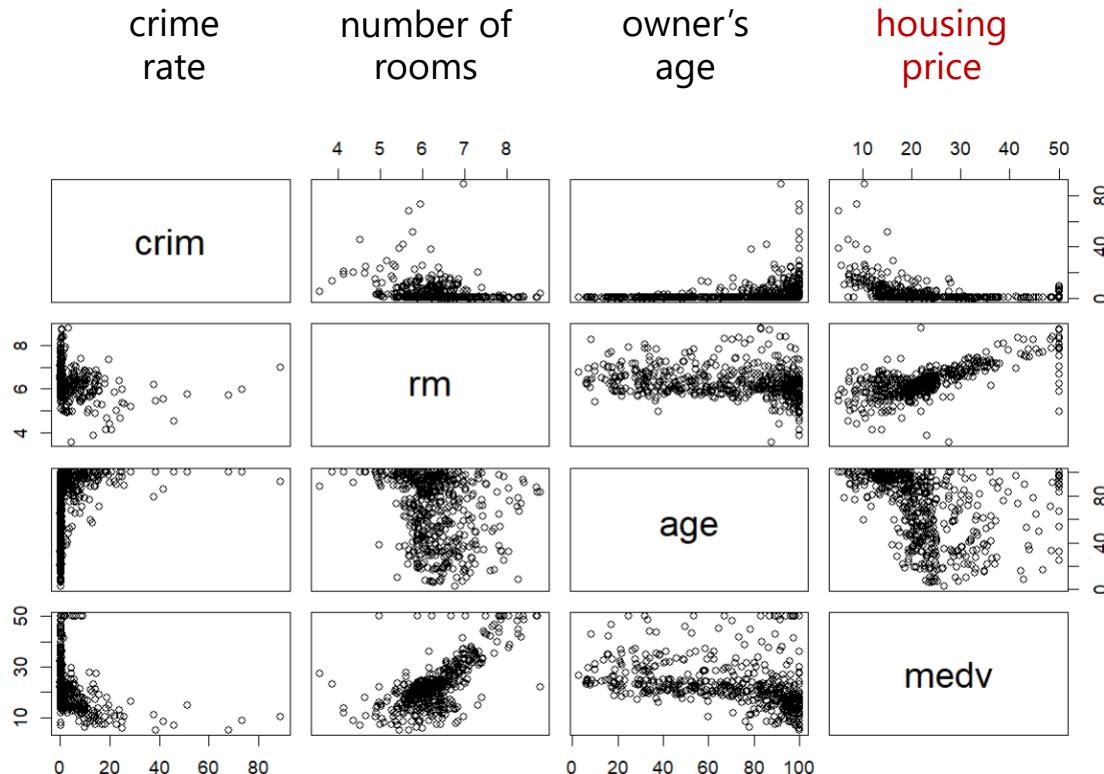
covariance

variance

|          | $X_1$    | $X_2$    | $\dots$  | $T$ |
|----------|----------|----------|----------|-----|
| $x_{11}$ | $x_{12}$ | $\dots$  | $x_{1p}$ |     |
| $x_{21}$ | $x_{22}$ | $\dots$  | $x_{2p}$ |     |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |     |
| $x_{n1}$ | $x_{n2}$ | $\dots$  | $x_{np}$ |     |
| $x_{1p}$ | $x_{2p}$ |          |          |     |

# Random Vector

# Variance-covariance matrix



```
```{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  |

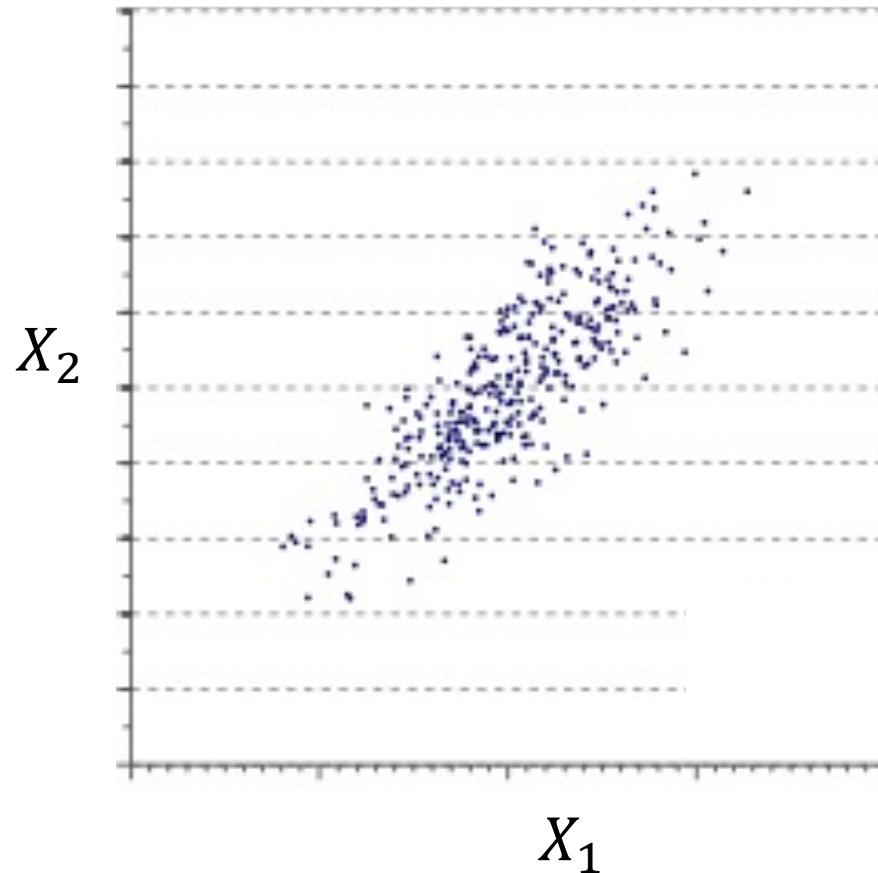
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

# Random Vector

## Exercise: the variance-covariance matrix

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

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



Q. What would be the correct  $\Sigma$  ?

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

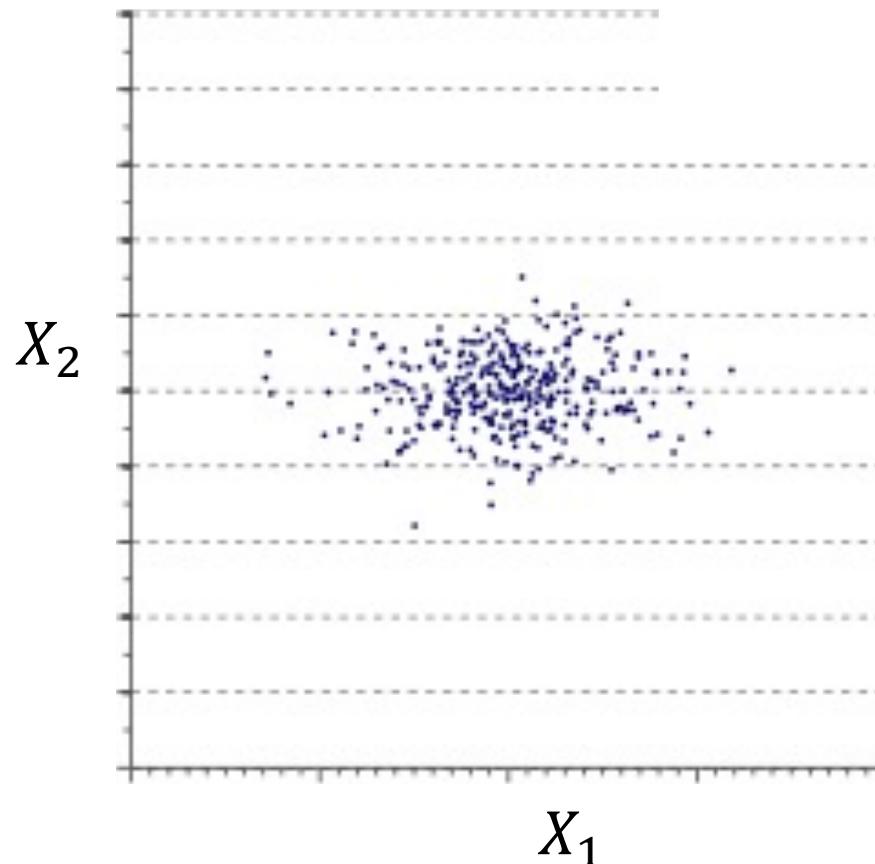
$$\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

# Random Vector

## Exercise: the variance-covariance matrix

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

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



Q. What would be the correct  $\Sigma$  ?

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

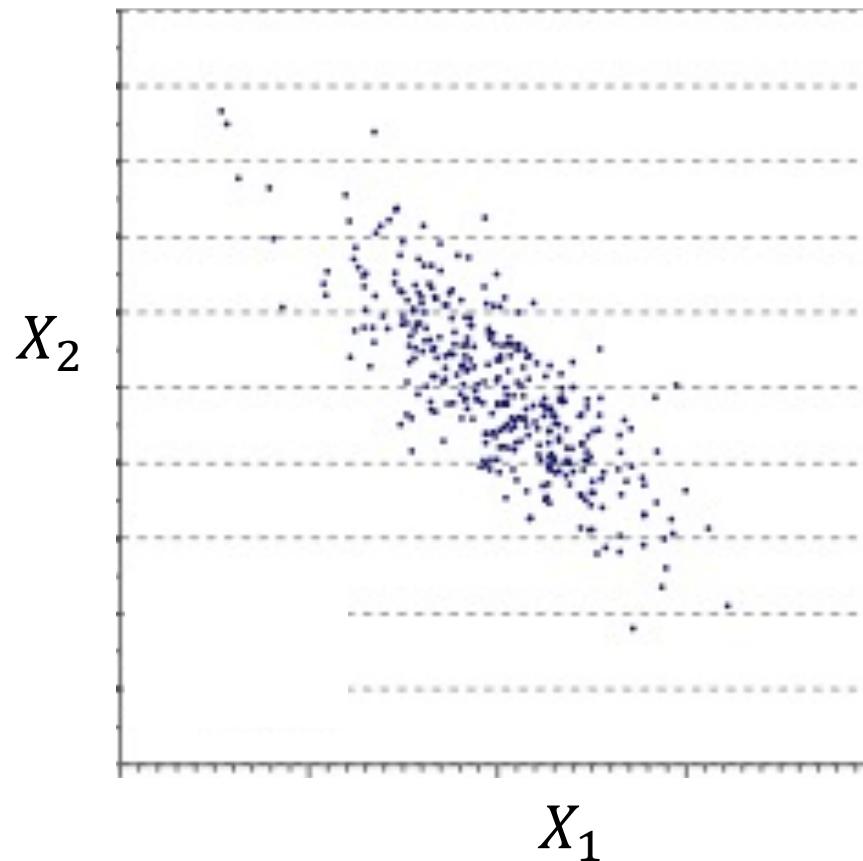
$$\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

# Random Vector

## Exercise: the variance-covariance matrix

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

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



Q. What would be the correct  $\Sigma$  ?

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

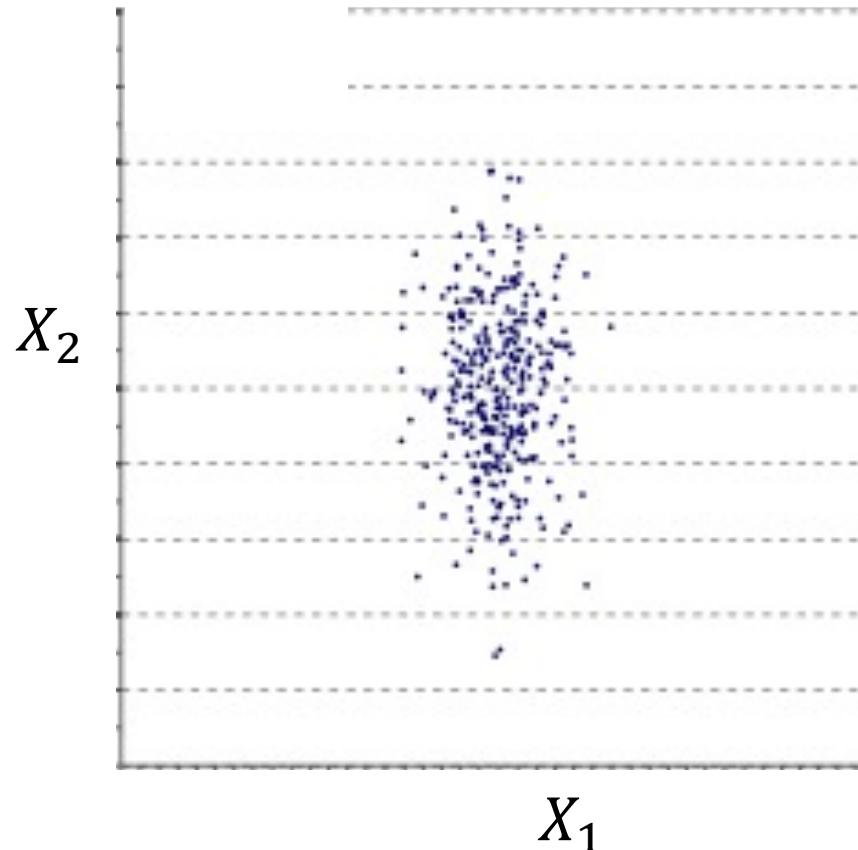
$$\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

# Random Vector

## Exercise: the variance-covariance matrix

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

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



Q. What would be the correct  $\Sigma$  ?

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  $\rho$  vs Covariance matrix  $\Sigma$

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

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

- $\rho$  is a normalized version of  $\Sigma$

# Random Vector

## Correlation matrix

Correlation matrix  $\rho$  vs Covariance matrix  $\Sigma$

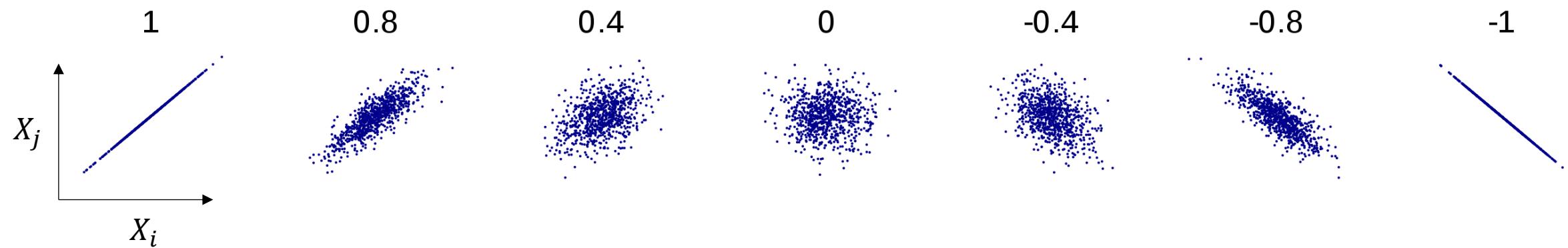
$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

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

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

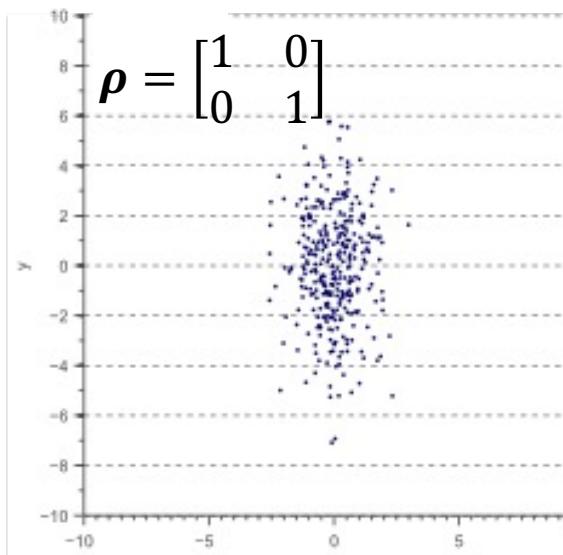
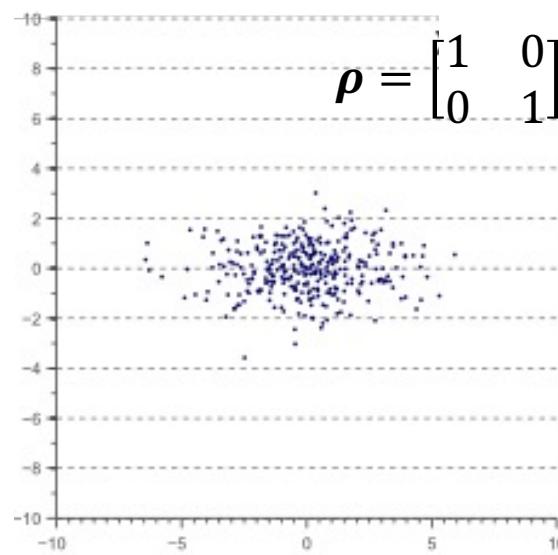
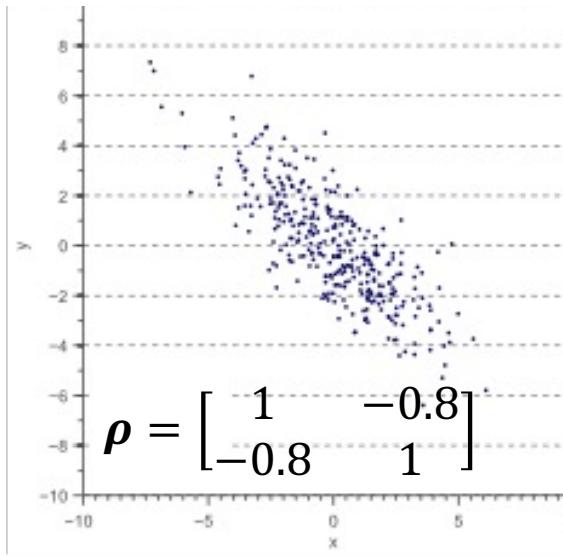
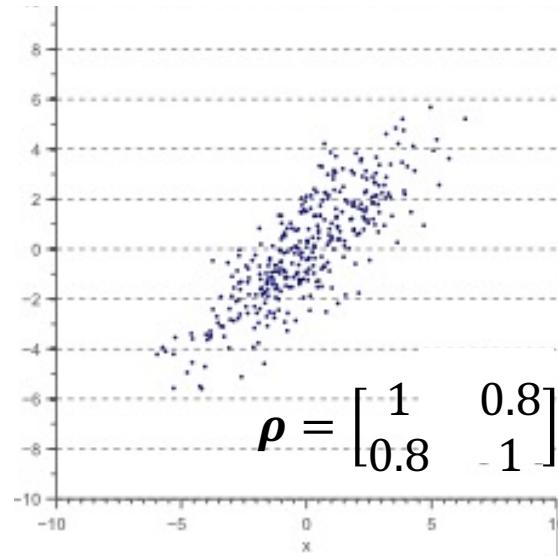
# Random Vector

How does  $\rho$  look?



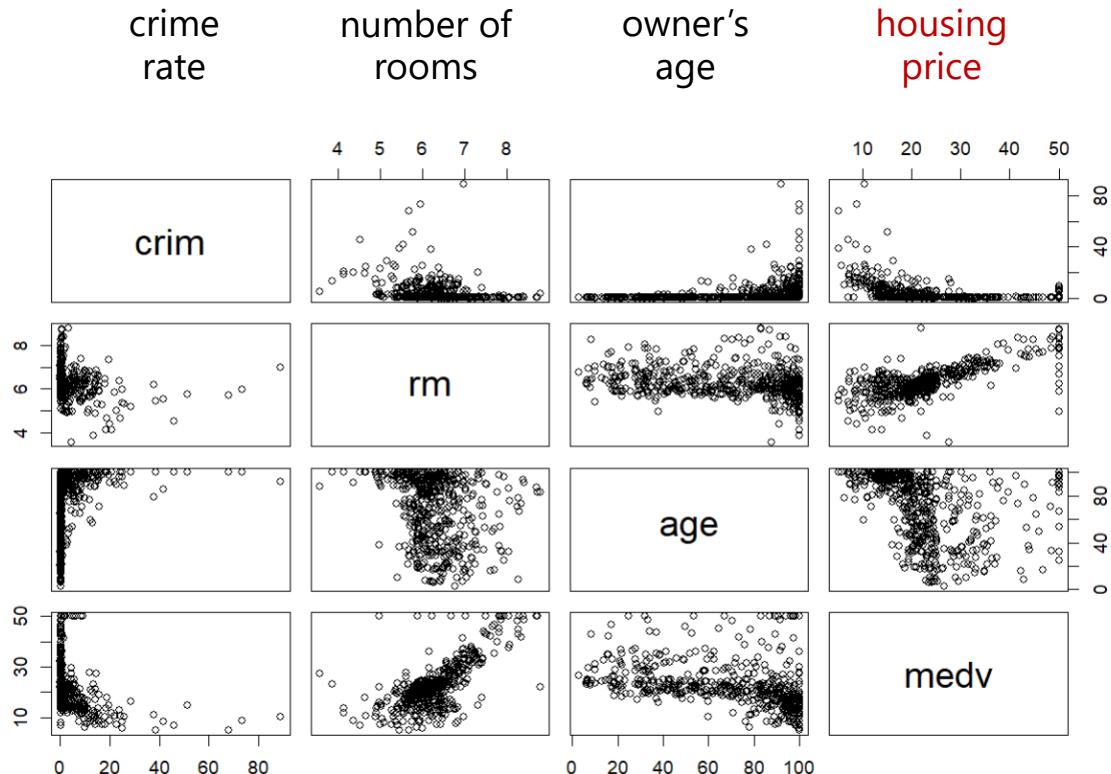
# Random Vector

How does  $\rho$  look?



# Random Vector

## Correlation matrix on the Boston Housing price dataset



```
```{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  |

## Linear combinations

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

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

Then,

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

### Proof of $Cov(\mathbf{Z})$

$$\begin{aligned} &= E[(\mathbf{Z} - E[\mathbf{Z}])(\mathbf{Z} - E[\mathbf{Z}])^T] \\ &= E[(\mathbf{C}\mathbf{X} - E[\mathbf{C}\mathbf{X}])(\mathbf{C}\mathbf{X} - E[\mathbf{C}\mathbf{X}])^T] \\ &= E[(\mathbf{C}\mathbf{X} - \mathbf{C}E[\mathbf{X}])(\mathbf{C}\mathbf{X} - \mathbf{C}E[\mathbf{X}])^T] \\ &= E[(\mathbf{C}\mathbf{X} - \mathbf{C}\mu_X)(\mathbf{C}\mathbf{X} - \mathbf{C}\mu_X)^T] \\ &= E\left[\mathbf{C}(\mathbf{X} - \mu_X)(\mathbf{C}(\mathbf{X} - \mu_X))^T\right] \\ &= E[\mathbf{C}(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T\mathbf{C}^T] \\ &= \mathbf{C}E[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T]\mathbf{C}^T \\ &= \mathbf{C}\Sigma_X\mathbf{C}^T \end{aligned}$$

# Random Vector

## Univariate case of $\text{Cov}(\mathbf{Z})$

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

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

Then,

$$\bullet \quad \text{Var}(c\mathbf{X}) = c^2 \text{Var}(\mathbf{X})$$

### Proof of $\text{Var}(c\mathbf{X})$

$$\text{Var}(\mathbf{X}) = E[(\mathbf{X} - E[\mathbf{X}])^2]$$

$$\begin{aligned}\text{Var}(c\mathbf{X}) &= E[(c\mathbf{X} - E[c\mathbf{X}])^2] \\ &= E[(c\mathbf{X} - cE[\mathbf{X}])^2]\end{aligned}$$

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

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

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

$$= c^2 \text{Var}(\mathbf{X})$$

# Random Vector

## Exercise: Linear combinations

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

|                             |
|-----------------------------|
| $X_1 - X_3$                 |
| $X_2 + X_4$                 |
| $(X_2 + X_4) - (X_1 + X_3)$ |

=

|          |          |          |          |
|----------|----------|----------|----------|
| $c_{11}$ | $c_{12}$ | $c_{13}$ | $c_{14}$ |
| $c_{21}$ | $c_{22}$ | $c_{23}$ | $c_{24}$ |
| $c_{31}$ | $c_{32}$ | $c_{33}$ | $c_{34}$ |

$Y$

$C$

|       |
|-------|
| $X_1$ |
| $X_2$ |
| $X_3$ |
| $X_4$ |

$X$

# Random Vector

## Exercise: Linear combinations

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

|                             |
|-----------------------------|
| $X_1 - X_3$                 |
| $X_2 + X_4$                 |
| $(X_2 + X_4) - (X_1 + X_3)$ |

=

|          |          |          |          |
|----------|----------|----------|----------|
| 1        | 0        | -1       | 0        |
| $c_{21}$ | $c_{22}$ | $c_{23}$ | $c_{24}$ |
| $c_{31}$ | $c_{32}$ | $c_{33}$ | $c_{34}$ |

|       |
|-------|
| $X_1$ |
| $X_2$ |
| $X_3$ |
| $X_4$ |

$Y$

$C$

$X$

# Random Vector

## Exercise: Linear combinations

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

$$\begin{bmatrix} X_1 - X_3 \\ X_2 + X_4 \\ (X_2 + X_4) - (X_1 + X_3) \end{bmatrix}$$

$Y$

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

$C$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$X$

Then,  $\text{Cov}(Y)$ ?

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

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

# Quiz

## Q1: Mean of sum

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

What is  $E(X + Y)$ ?

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

## Q2: Mean of linear combination

Q.  $X$  is a 2-dimensional random vector with  $E[X] = (2, 5)^T$ , and  $b = (0.5, 0.5)^T$  is a constant vector. What is  $E[b^T X]$ ?

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

## Q3: Covariance

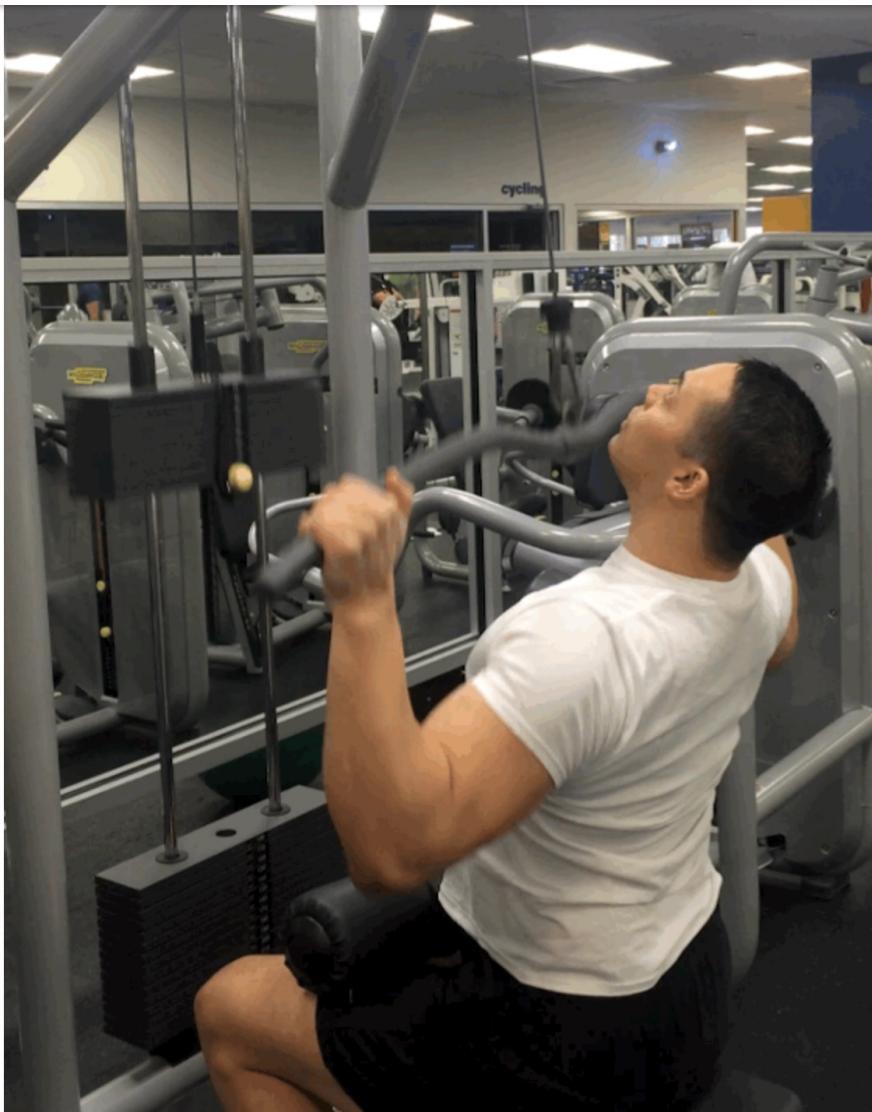
Q.  $X$  is a  $p$ -dimensional random vector with mean  $\mu$ , and  $X \in \mathbb{R}^{p \times n}$ . Which of the following defines the covariance matrix?

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

# Multivarite Normal Distribution

# Multivariate Normal (mvN) Distribution

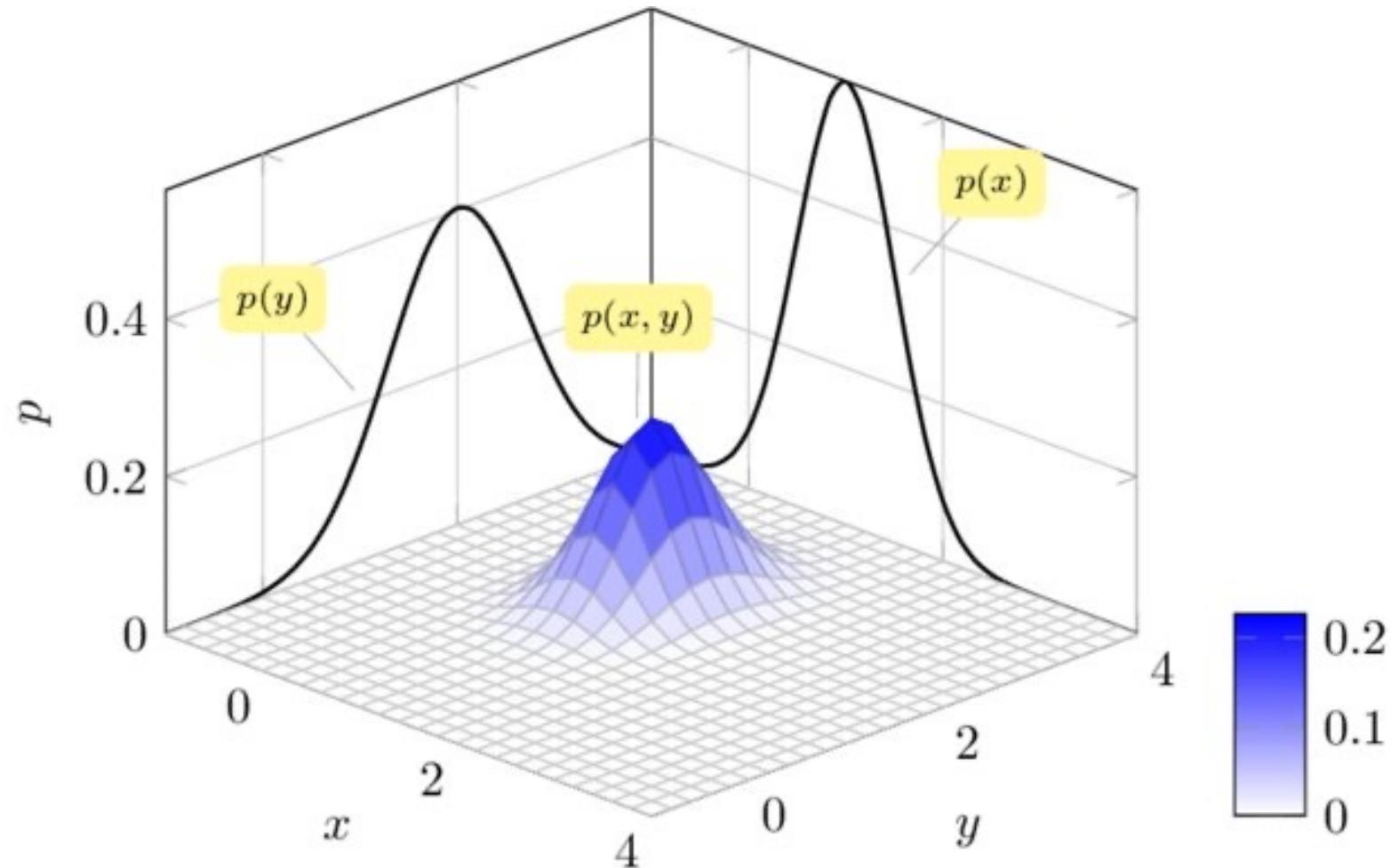
## Why is the mvN distribution so popular?



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

# Multivariate Normal Distribution

## Multivariate Normal (mvN) Distribution



# Multivariate Normal Distribution

## Multivariate Normal Probability Density Function (PDF)

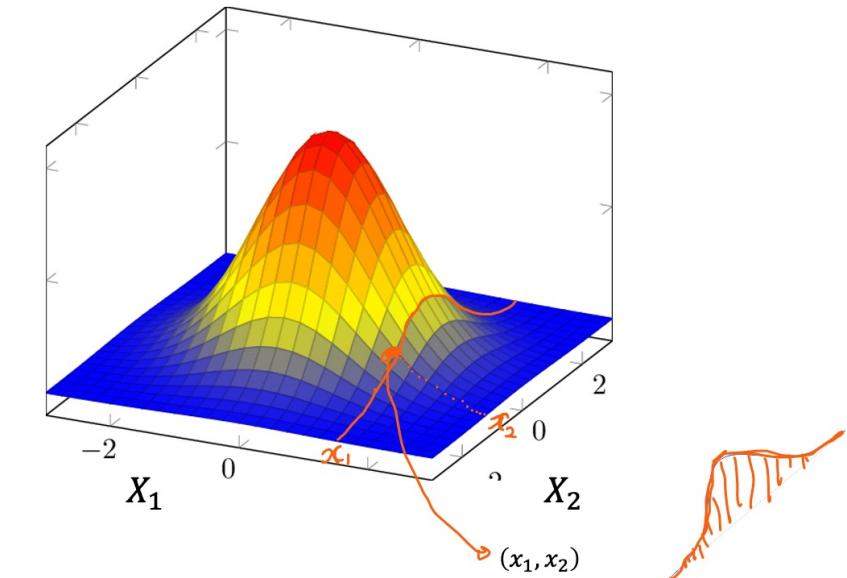
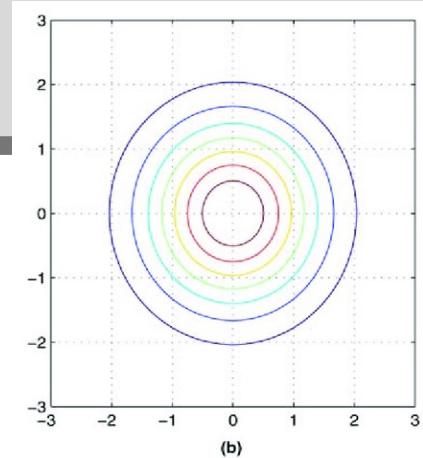
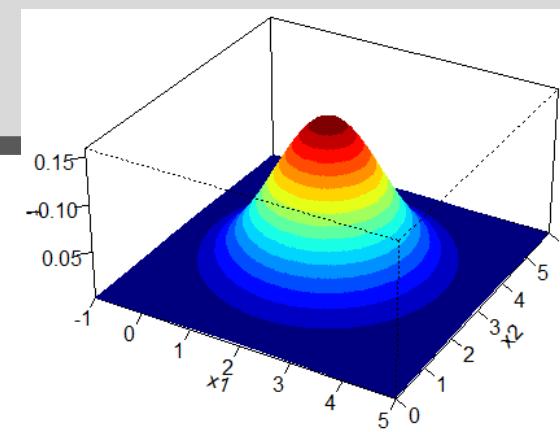
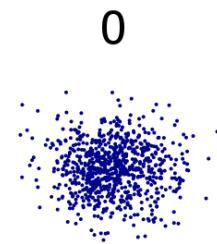
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

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

# Multivariate Normal Distribution

## Useful properties of the mvN distribution

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



# Multivariate Normal Distribution

## **Some note**

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

# Multivariate Normal Distribution

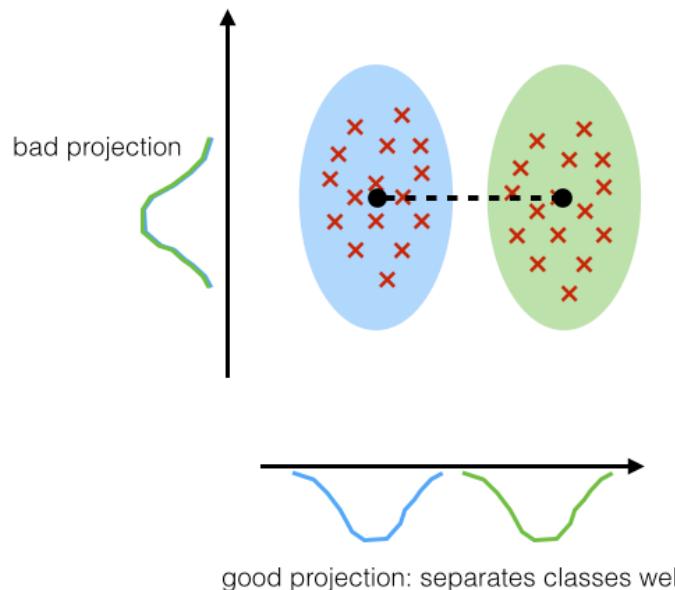
## Wait! Why are we learning it?

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

### LDA:

maximizing the component axes for class-separation

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

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right\}$$

Q. What is  ?

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

## Question 2: Multivariate normal distribution

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

# Multivariate Normal Distribution

## Question 3: Independence

Let  $\mathbf{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , 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  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . How can I construct a vector of independent standard normal variables from  $\mathbf{X}$ ?

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

A group of six children, three boys and three girls, are seen from behind running down a school hallway. They are all wearing backpacks and casual clothing. The hallway has white walls and doors on either side. The children are in various stages of motion, with some arms raised. The lighting is bright, typical of an indoor school environment.

The End!