데이터사이언스응용
(Capstone design)

김응희

ehkim@sunmoon.ac.kr

Week 04

Notice #1


SW중심대학
2020 하반기 특별평가 및 접수안내
TOPCIT
응시료
10,000원
응시여부에 따라 장학금(30,000원) 지급예정
접수방법
① TOPCIT 홈페이지(http://www.topcit.or.kr/) 접속 후 아이디 생성
② 첨부파일-신청서 작성 후, 이메일(swedu@sunmoon.ac.kr)로 제출
③ 응시료 10,000원 아래와 같이 입금
-입금 계좌: 국민)273802-04-177278 주현진
-입금자명: 본인학번 기재
* 응시료 입금 확인이 되지 않은 경우, 접수 불가
접수기간
20.9.21.(월) 09:00~ 9.28(월) 17:00
평가일시
20.10.31(토) 09:30~12:00 *09:10까지 입실 완료
평가장소
선문대학교 원화관 실습실 (자세한 장소는 추후공지)
혜 택
성적 우수자 학생 장학금 지급
☎ 문의 : SW교육지원센터 041-530-8310
※ 위 일정은 코로나19 관련 상황 및 정부 지침에 따라 일정 변공(연기 또는 취소) 될 수 있음

Notice #1


SW중심대학
2020 하반기 특별평가
TOPCIT
온라인 설명회 안내
일시: 2020.09.23(수) 13:30~14:30
장소: 온라인 ZOOM
* URL- https://zoom.us/j/3937852058
문의 : SW교육지원센터 041)530-8310
혜 택
성적 우수자 학생 장학금 지급
상위 1등: 50만 원 / 상위 2등: 30만원

Notice #1

- e-강의동 > 데이터사이언스응용 > 과제 > 과제 0. TOPCIT 신청서
  - 첨부파일 다운로드 후, 작성
  - 파일 이름: 2020_하반기_TOPCIT_특별평가_신청서양식_이름_학번.hwp
  - 제출 기한: 2020.09.28. 15:00

NOTICE #2

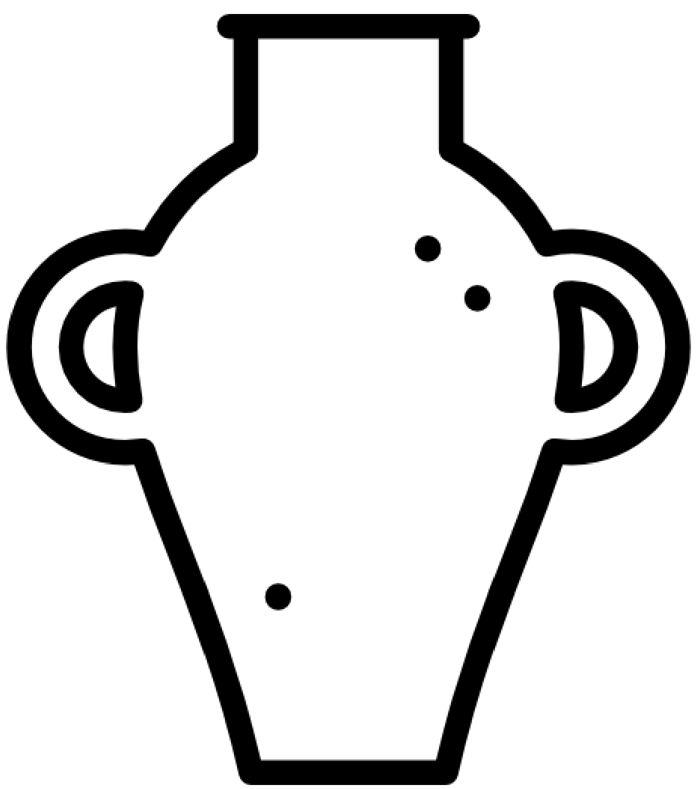
순번	팀명	팀원	주제
1	제니리아	쿠마자와 유이 후쿠미쓰 치아키	Best eyebrow
2	Ajsoftware	노승욱 스피겔 크릴	I2M Generating music through image input
3	H:J	이양희 이수정 이혜인	?
4	안시성	우메모토세이야 방대호 노무라 타카미치	냉장고를 부탁해 보유 식자재 기반 메뉴 추천 AI
5	철딱써니	정철우 김선민	눈부신 순간들 Bright moments
6	AKI	호즈미요시아키 오타오아키	Image arranger
7	YOLO	유제훈 키타야마요시야키	CA-forecaster Car accident forecaster

MLE

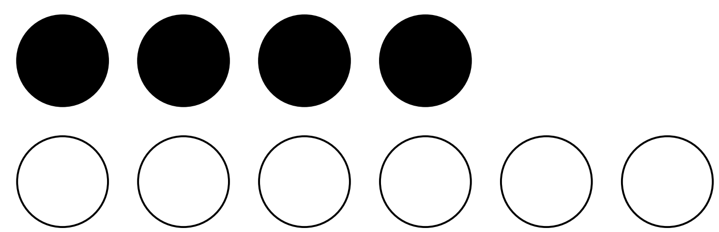
(Maximal likelihood estimation)

Quiz

of stones: 100



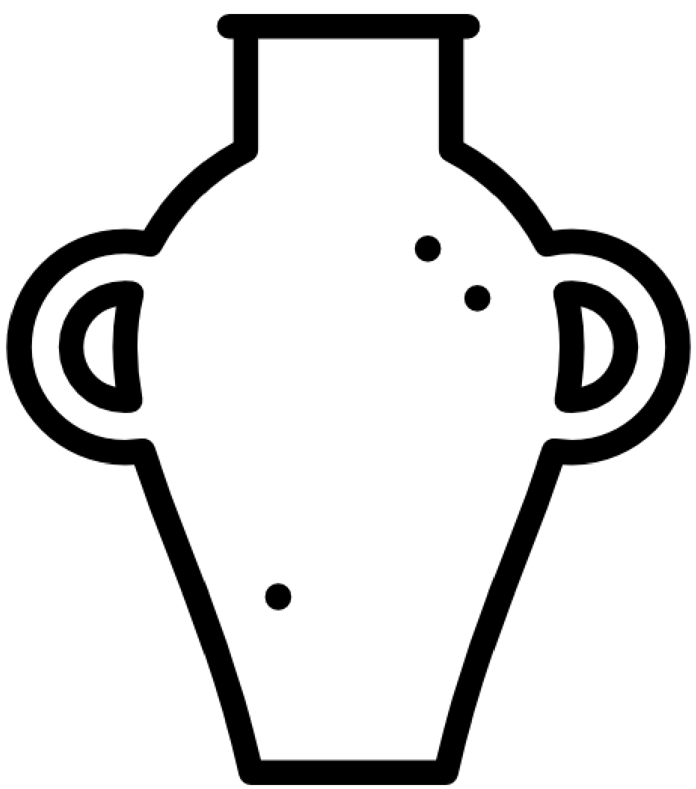
Result of sampling 10 times with replacement



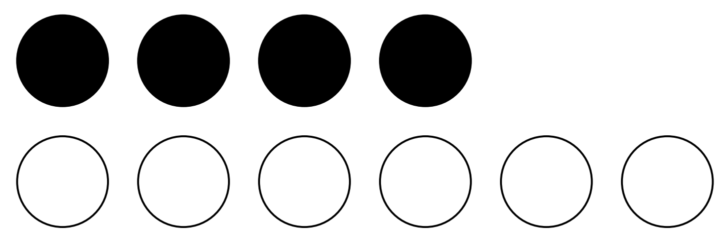
How many are the **black stones** in the pottery?

Quiz

of stones: 100



Result of sampling 10 times with replacement

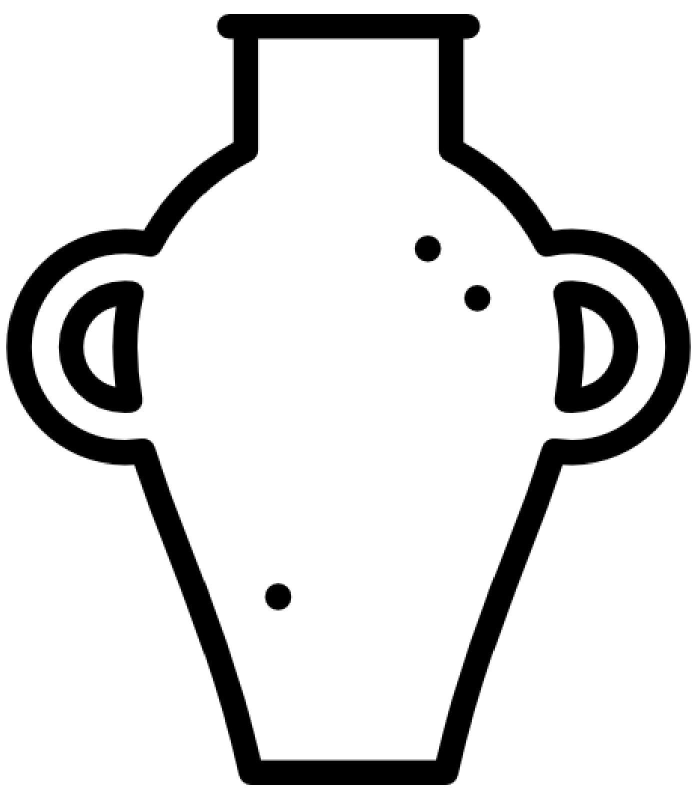


How many are the **black stones** in the pottery?

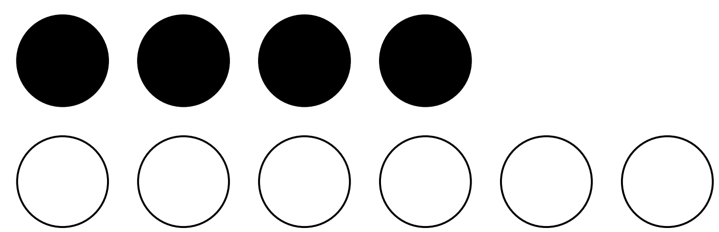
40

Quiz

of stones: 100



Result of sampling 10 times with replacement

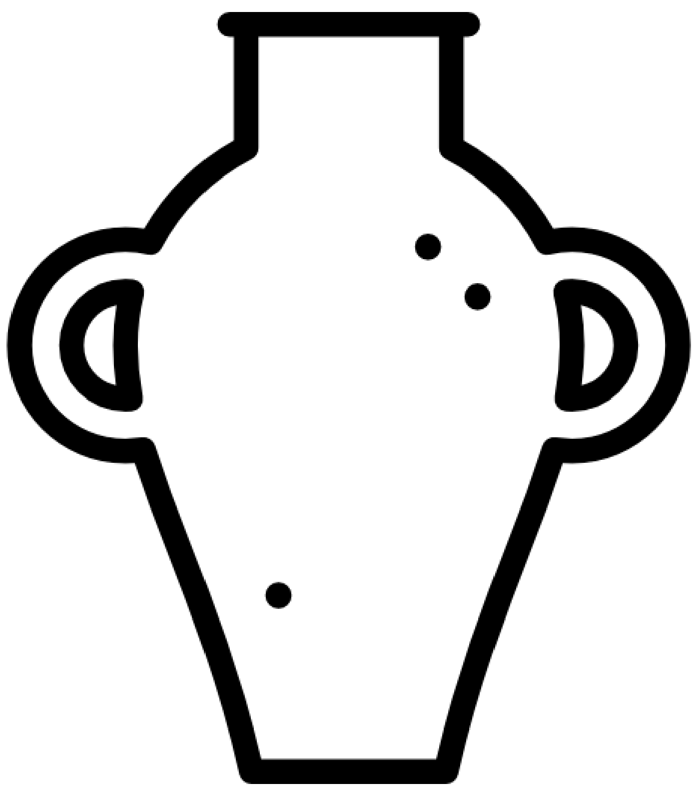


How many are the **black stones** in the pottery?

40

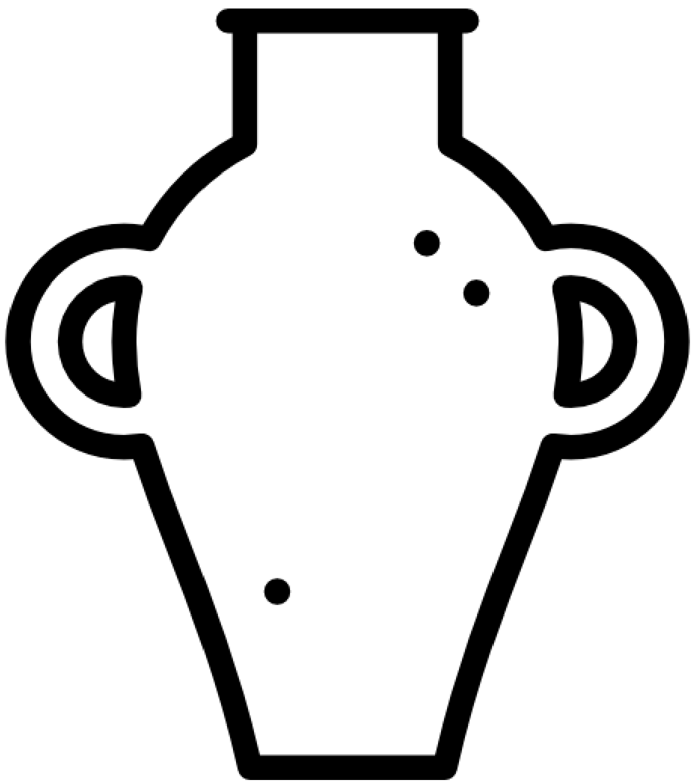
Could you explain why?

What we want to know is p



$$p = \frac{\#\ of\ black\ stones}{\#\ of\ stones}$$

What we want to know is p



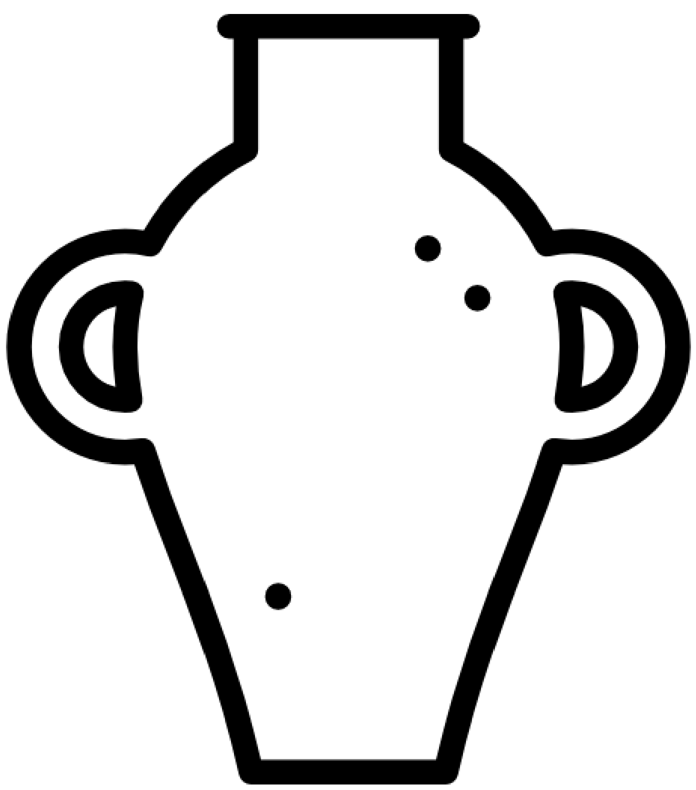
$$p = \frac{\#\ of\ black\ stones}{\#\ of\ stones}$$

As a by–product

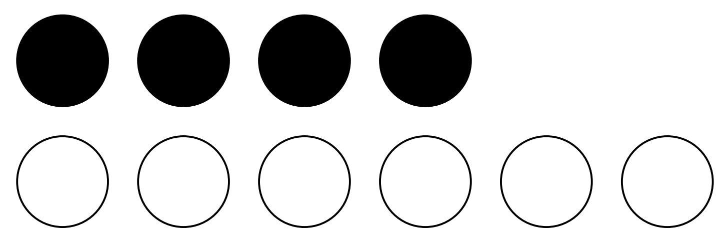
$$1 - p = \frac{\#\ of\ white\ stones}{\#\ of\ stones}$$

Can we consider our observation as a **probabilistic event**?

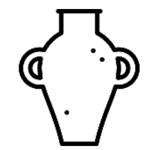
# of stones: 100



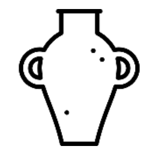
Result of sampling 10 times with replacement



Can we consider our observation as a probabilistic event?

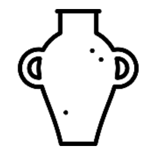
$P($ ●○○●○●○●○○ $|$  $)$

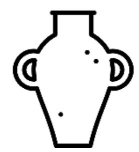
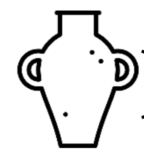
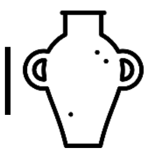
Can we consider our observation as a probabilistic event?

$P($ ●○○●○●○●○○ $|$  $)$

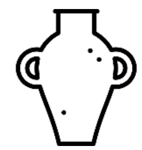
Result of sampling 10 times with replacement → The event, sampling each stone, is **independent** to each other.

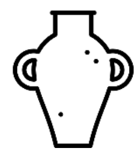
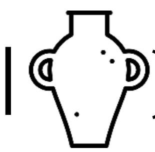
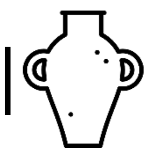
Can we consider our observation as a probabilistic event?

$P(\bullet\circ\circ\bullet\circ\bullet\circ\bullet\circ\circ \mid$  $)$

$= P(\bullet \mid$  $) \times P(\circ \mid$  $) \times \cdots \times P(\circ \mid$  $)$

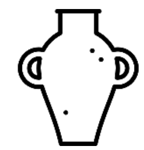
Can we consider our observation as a probabilistic event?

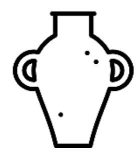
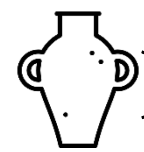
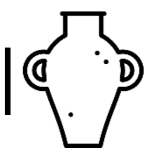
$P(\bullet\circ\circ\bullet\circ\bullet\circ\bullet\circ\circ \mid$  $)$

$= P(\bullet \mid$  $) \times P(\circ \mid$  $) \times \cdots \times P(\circ \mid$  $)$

$$p = \frac{\#\ of\ black\ stones}{\#\ of\ stones} \qquad 1 - p = \frac{\#\ of\ white\ stones}{\#\ of\ stones}$$

Can we consider our observation as a probabilistic event?

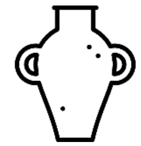
$P(\bullet \circ \circ \bullet \circ \bullet \circ \bullet \circ \circ \mid$  $)$

$= P(\bullet \mid$  $) \times P(\circ \mid$  $) \times \cdots \times P(\circ \mid$  $)$

$$= p \times (1 - p) \times \cdots \times (1 - p)$$

$$= p^4 \times (1 - p)^6$$

Can we consider our observation as a probabilistic event?

$P($●○○●○●○●○○ $\mid$ $) = p^4 \times (1-p)^6$

Other cases

●●●●○○○○○○

●●●○●○○○○○

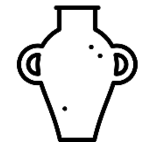
●●●○○●○○○○

...

○○○○○○●●●●

How many case are there?

Can we consider our observation as a probabilistic event?

$P($ ●○○●○●○●○○ $|$  $) = p^4 \times (1-p)^6$

Other cases

●●●●○○○○○○

●●●○●○○○○○

●●●○○●○○○○

...

○○○○○○●●●●

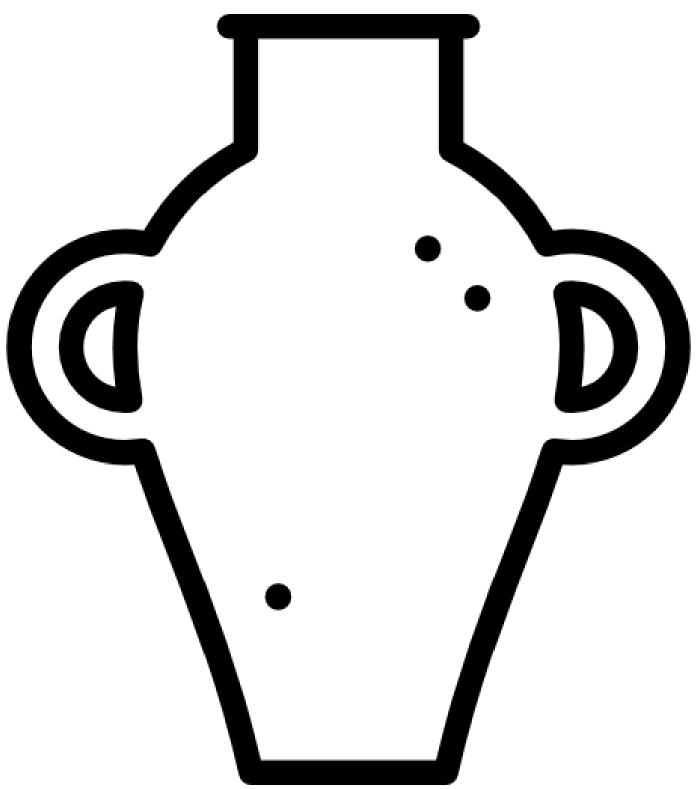
How many case are there?

01	02	03	04	05	06	07	08	09	10

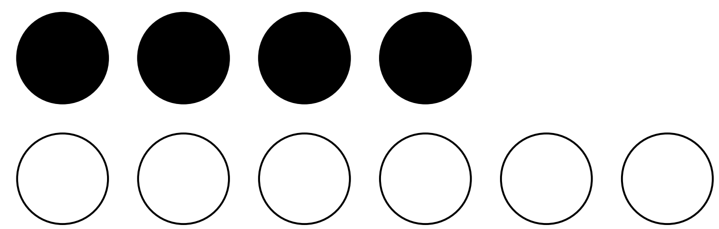
$${}_{10}C_4 \text{ w/o repetition} = \binom{10}{4} = \frac{10!}{4!(10-4)!} = 210$$

Can we consider our observation as a probabilistic event?

# of stones: 100



Result of sampling 10 times with replacement

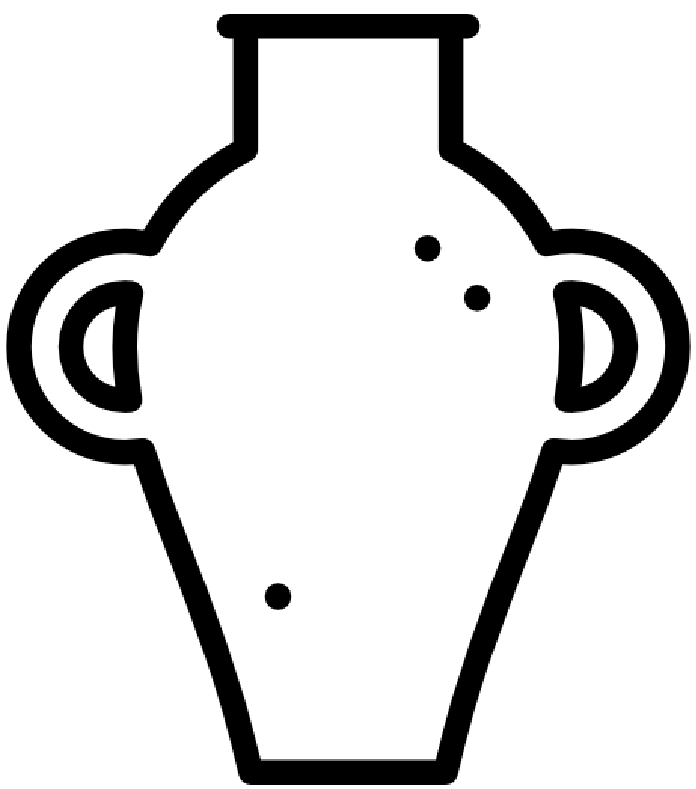


Probability of our observation

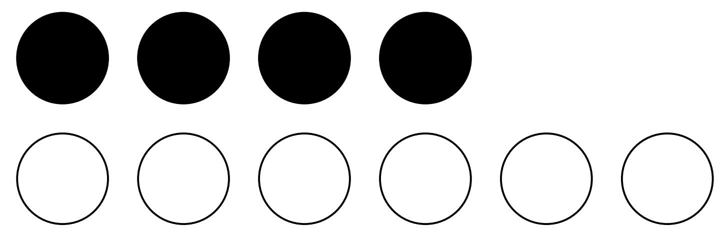
$$210 \times p^4 \times (1-p)^6$$

Can we consider our observation as a **probabilistic event**?

# of stones: 100



Result of sampling 10 times with replacement



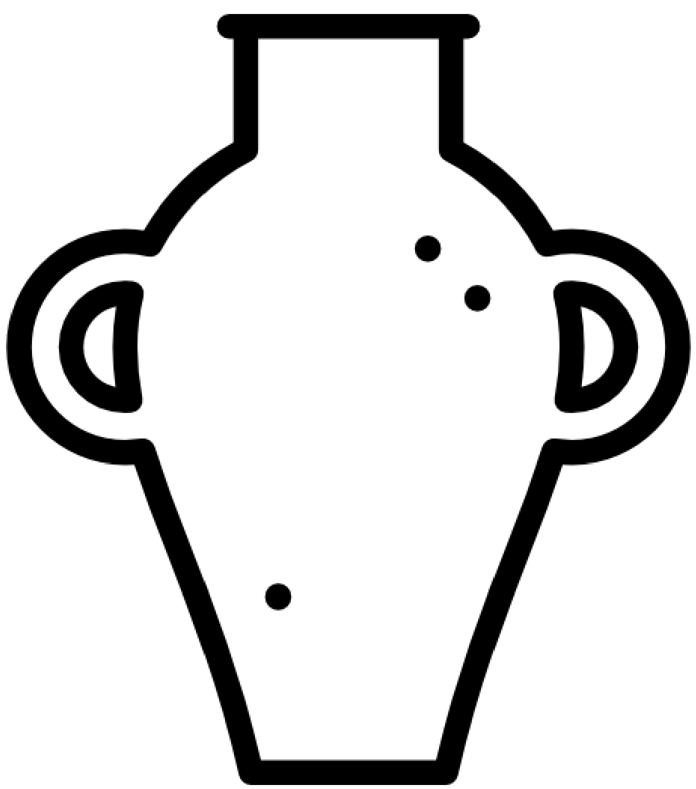
Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

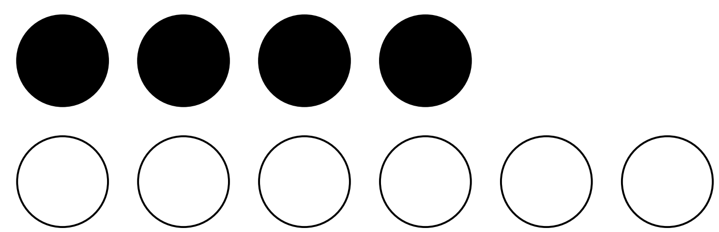
What should we do from now on?

Can we consider our observation as a probabilistic event?

# of stones: 100



Result of sampling 10 times with replacement



Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

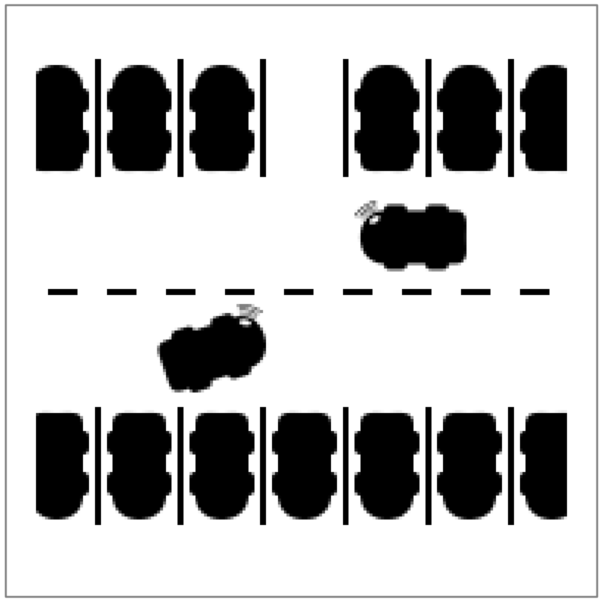
What should we do from now on?

We need to find the value p which **maximizes** the probability of our observation.

Why we need to maximize the probability of our observation

Our observation

Common?



Uncommon?

Why we need to maximize the probability of our observation

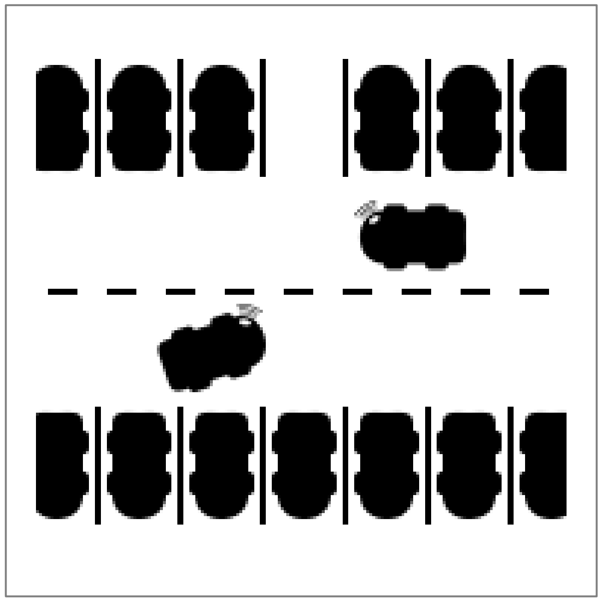
Our observation

Common

Uncommon?

Why we need to maximize the probability of our observation

Our observation



Common

↓

High frequency

↓

High probability

Uncommon?

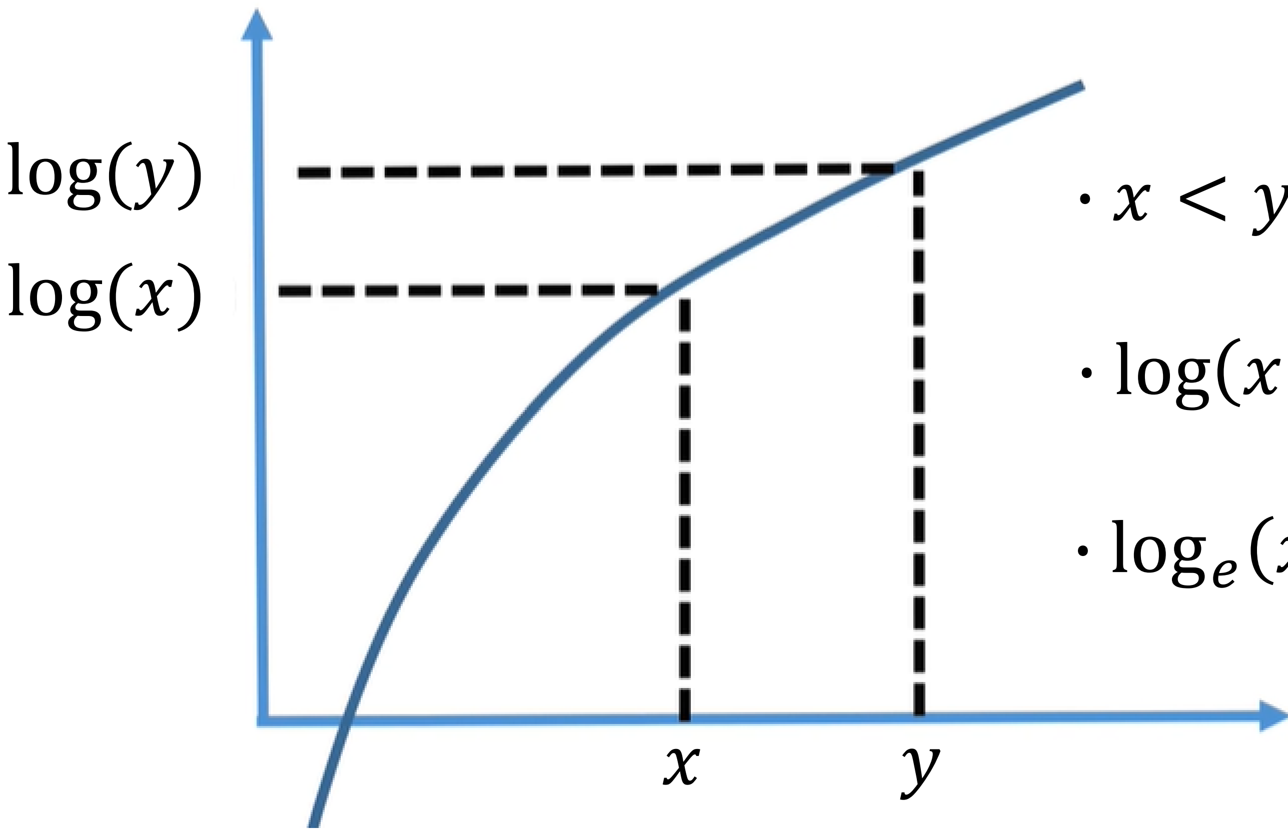
How to maximize the probability of our observation

Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p \, 210 \times p^4 \times (1-p)^6$$

How to maximize the probability of our observation



- $x < y \implies \log(x) < \log(y)$
- $\log(x \times y \times z) = \log(x) + \log(y) + \log(z)$
- $\log_e(x) = \ln(x) \longrightarrow (\ln(x))' = \dfrac{1}{x}$

How to maximize the probability of our observation

Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p\, 210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p\, \ln(210 \times p^4 \times (1-p)^6)$$

How to maximize the probability of our observation

Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p\, 210 \times p^4 \times (1-p)^6$$

$$\downarrow$$

$$p^* = \arg max_p\, \ln(210 \times p^4 \times (1-p)^6)$$

$$f(p) = \ln(210 \times p^4 \times (1-p)^6)$$

$$\downarrow$$

$$f(p) = \ln(210) + \ln(p^4) + \ln((1-p)^6)$$

$$\downarrow$$

$$f(p) = \ln(210) + 4\ln(p) + 6\ln(1-p)$$

How to maximize the probability of our observation

Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$$

$$f(p) = \ln(210) + 4\ln(p) + 6\ln(1-p)$$

How to maximize the probability of our observation

Probability of our observation

$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$$

$$f(p) = \ln(210) + 4\ln(p) + 6\ln(1-p)$$

$$f'(p) = \frac{4}{p} - \frac{6}{1-p}$$

← Give me a second
(Chain rule)

$$f'(p) = \frac{4}{p} - \frac{6}{1-p} = 0$$

How to maximize the probability of our observation

Probability of our observation

$$\boxed{210 \times p^4 \times (1-p)^6}$$

$$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$$

$$f(p) = \ln(210) + 4\ln(p) + 6\ln(1-p)$$

$$f'(p) = \frac{4}{p} - \frac{6}{1-p} = 0$$

How to maximize the probability of our observation

Probability of our observation
$$210 \times p^4 \times (1-p)^6$$

$$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$$

$$f(p) = \ln(210) + 4\ln(p) + 6\ln(1-p)$$

$$f'(p) = \frac{4}{p} - \frac{6}{1-p} = 0$$

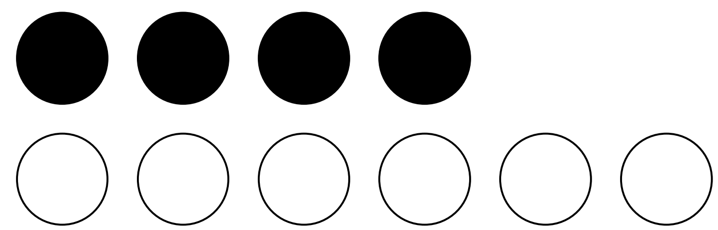
$$\frac{4}{p} = \frac{6}{1-p} \longrightarrow 4 - 4p = 6p \longrightarrow 4 = 10p \longrightarrow p = \frac{4}{10}$$

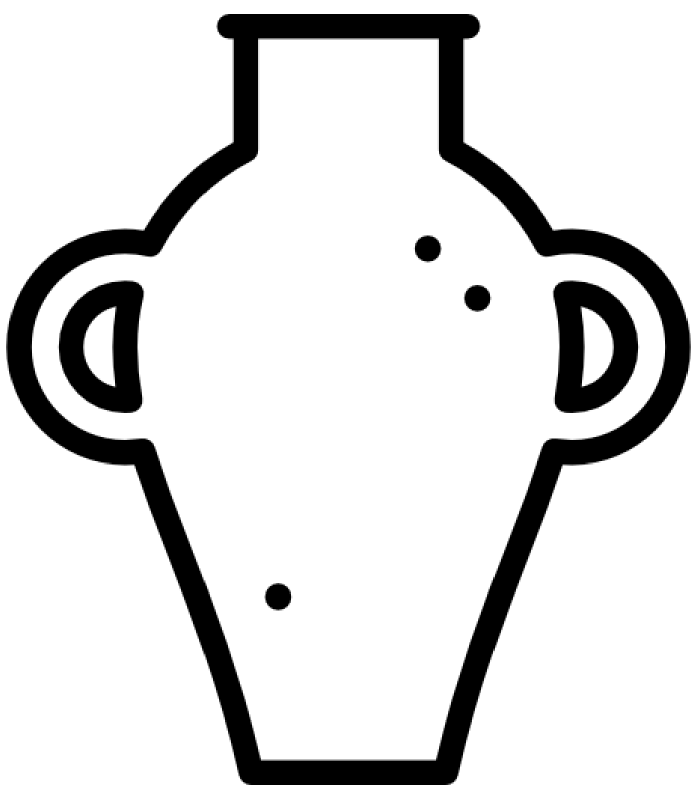
$$\therefore\ p^* = 0.4$$

Quiz

Result of sampling 10 times with replacement

# of stones: 100





$$p = \frac{\#\ of\ black\ stones}{\#\ of\ stones}$$

How many are the **<u>black stones</u>** in the pottery?

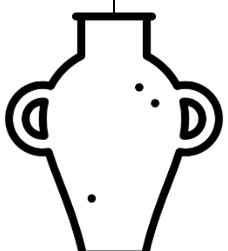
40

Terminology

- Likelihood
  - The hypothetical probability that an event that has already occurred would yield a specific outcome.
    - Hypothesis: p (proportion of black stones in the pottery)
    - Event already occurred: the result of 10 sampling (4 black stones & 6 white stones)
    - Hypothetical probability: probability of ●●●●○○○○○○ on  using the hypothesis p
    - Notation $(\theta \mid \vec{x})$: $L(p \mid \vec{x})$

- Maximum likelihood estimation (MLE)
  - the procedure of finding the value of one or more parameters for a given statistic which makes the known likelihood distribution a maximum
  - Result of Maximum likelihood estimation (MLE): $p = 0.4$
  - Assumption: we are not special → our observation is common → it should have as **high probability value** as possible.

Probability vs. Likelihood



$\vec{x}$

●●●●
○○○○○○

●●●●●
○○○○○

●●●●●●
○○○○

...

Probability
with fixed (known) parameter(s)
e.g., $p = 0.4$

$$P(\vec{x} \mid p) = \binom{10}{x} 0.4^3 (1 - 0.4)^{10-x}$$

Likelihood
with fixed (observed) data
e.g., $\vec{x}$: ●●●●
○○○○○○

p
0.1
0.2
...

$$L(p \mid \vec{x}) = \binom{10}{3} p^3 (1 - p)^{10-3}$$

MLE: Maximal likelihood estimation

This is why people say that

MLE is the basic & core technique of

the field of pattern recognition including

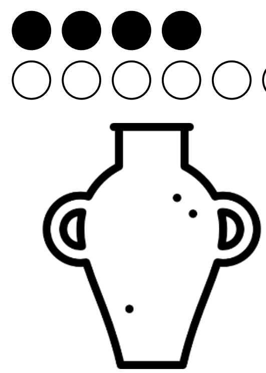
Deep learning, Support Vector Machine, Decision Trees,

Markov Random Field, Neural Networks, Linear Regression,

Logistic Regression, Maximum Entropy Model and etc.

Quiz summary

an event occurred

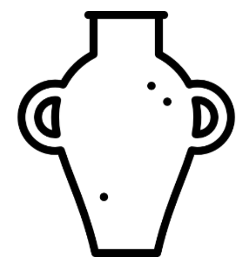


p

Proportion of black stones

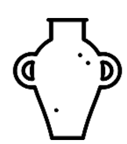
$1 - p$

Proportion of white stones



Finding p maximizing the probability of the event

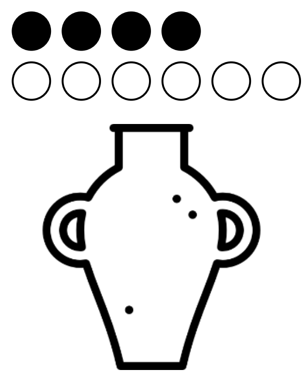
Using p that we found, now we can evaluate the probability of any event on



Labeling the process for quiz solving

Data
(Observation)

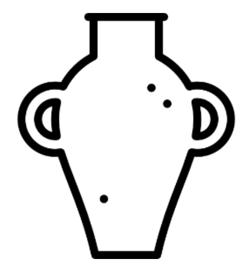
an event occurred



Model parameters

p
Proportion of black stones

$1 - p$
Proportion of white stones



Finding p maximizing the probability of the event

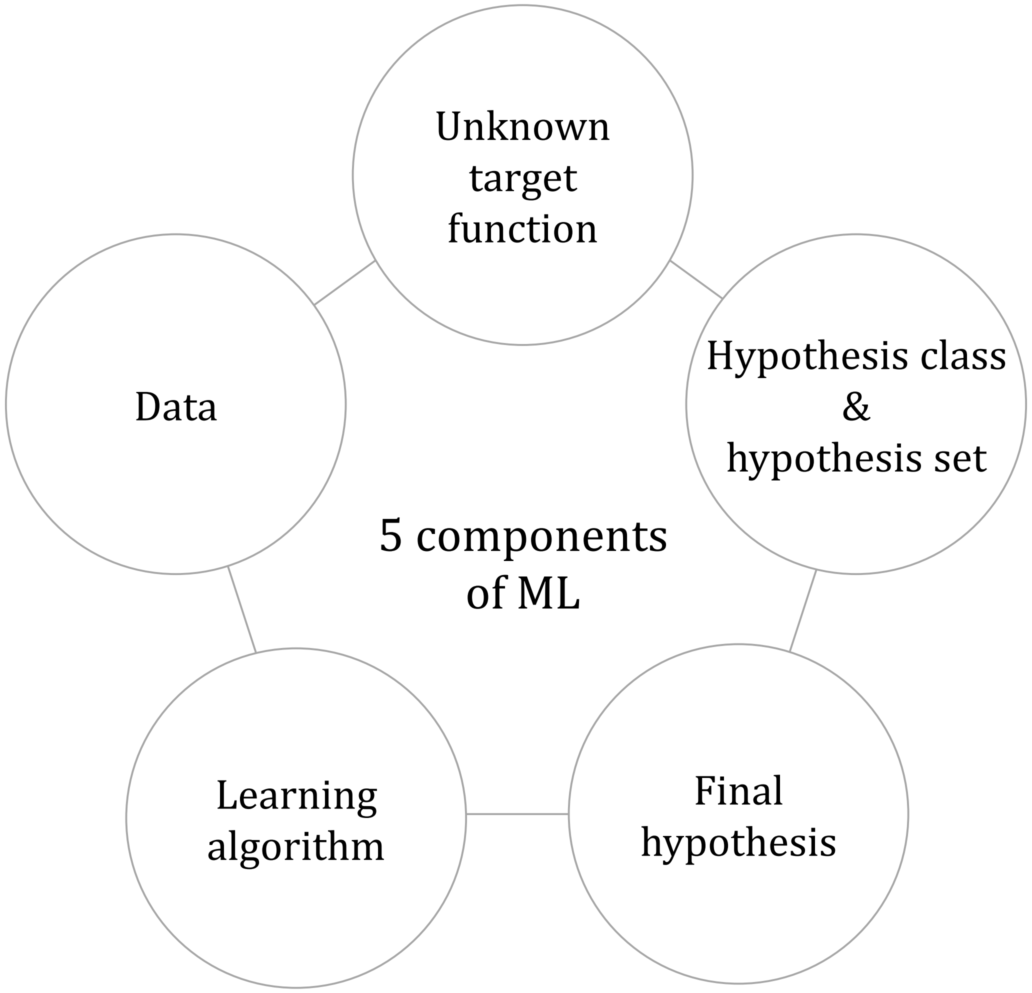
Learning

Using p that we found, now we can evaluate the probability of any event on

Using p that we found, now we can evaluate the probability of any event on

Prediction

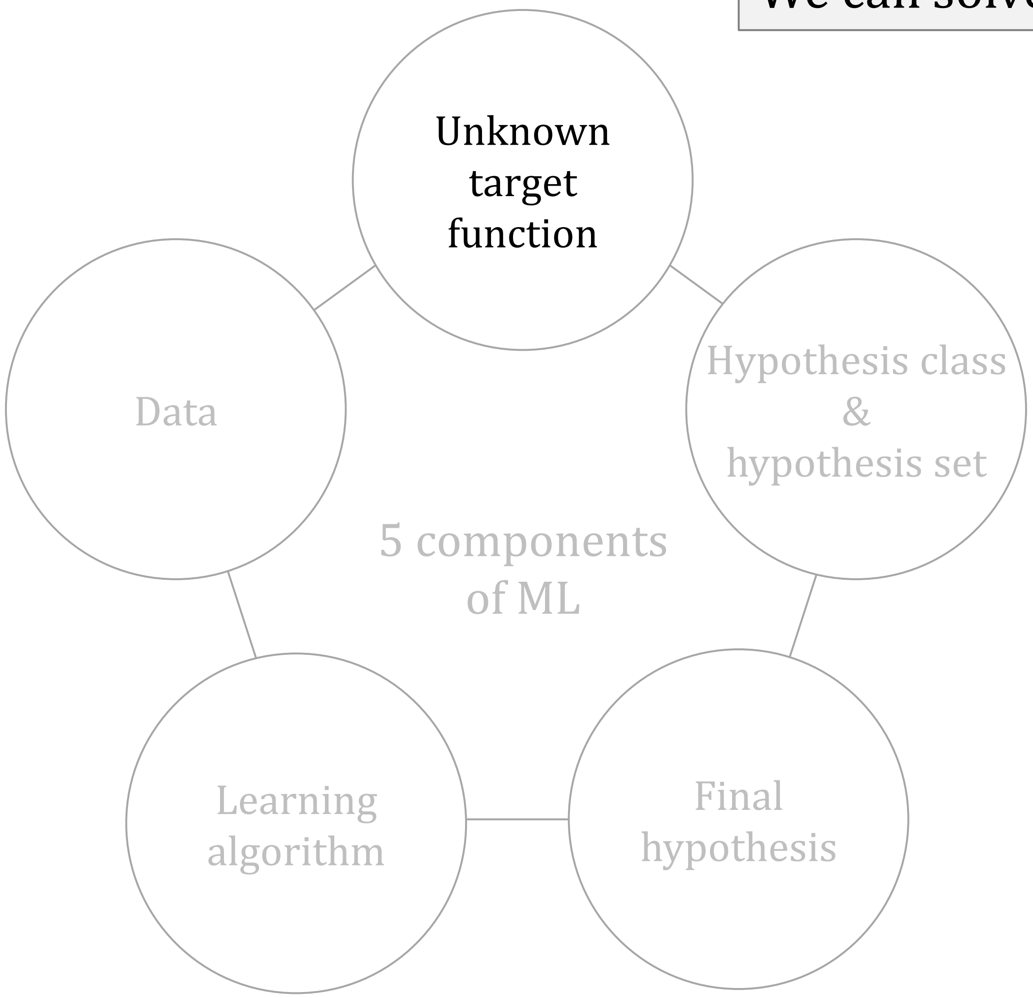
Do you remember?



Do you remember?

$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

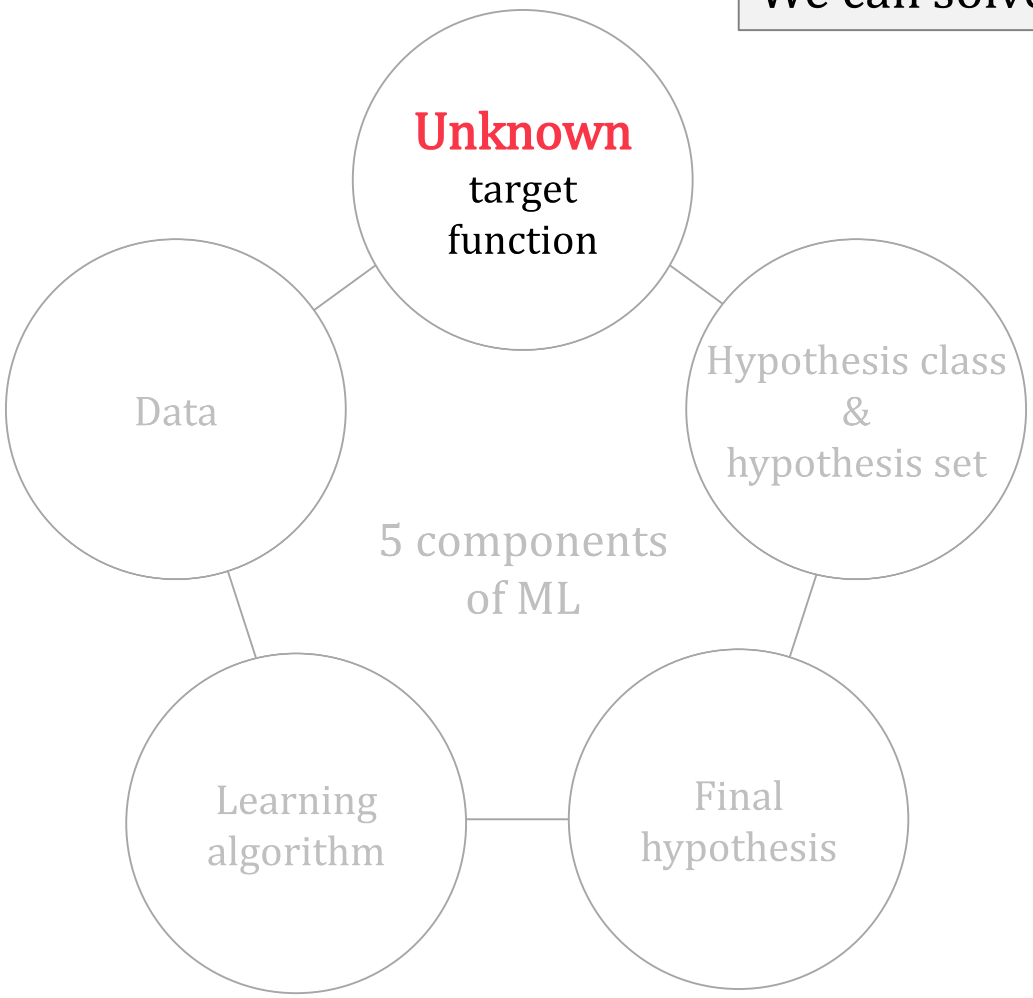
We can solve every problem!



Do you remember?

$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

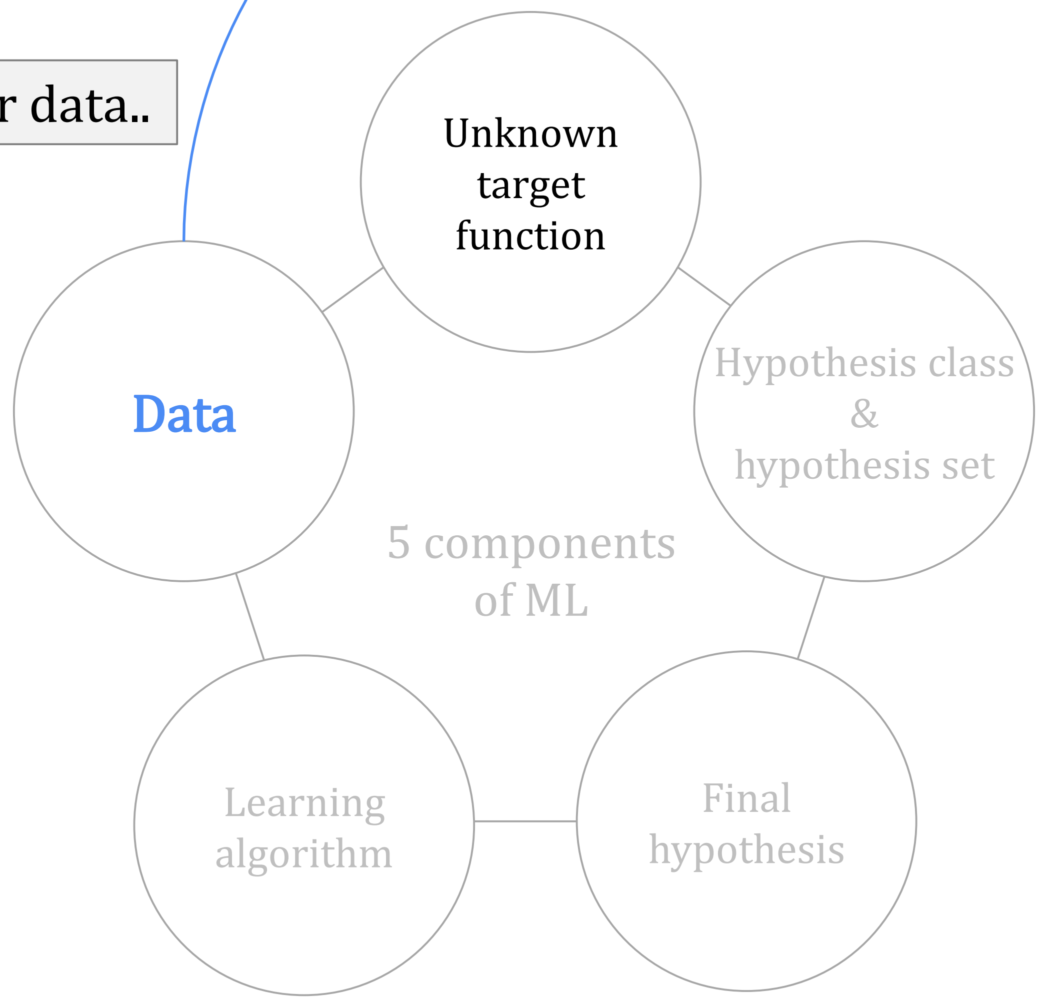
We can solve every problem!



Do you remember?

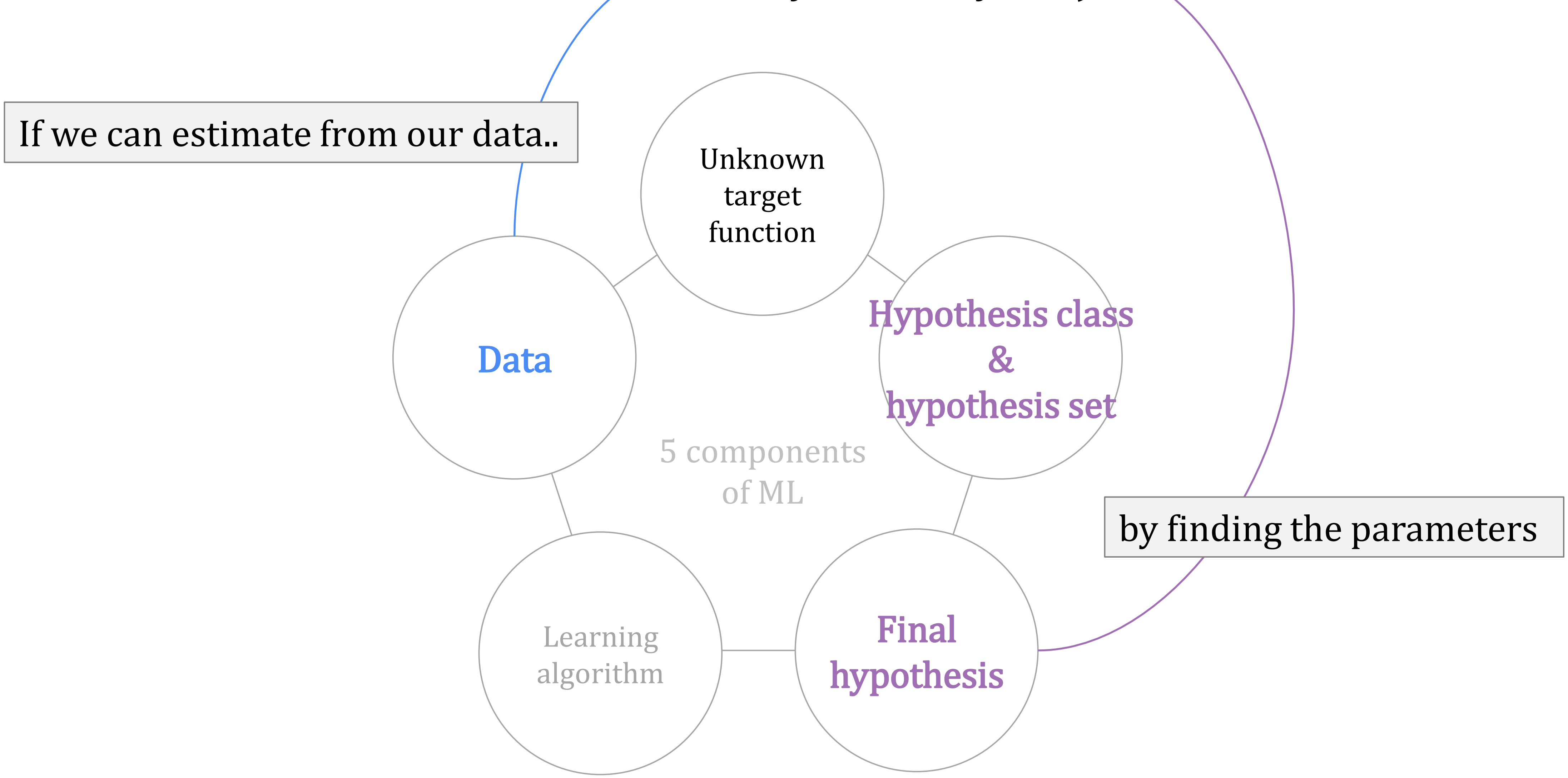
$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

If we can estimate from our data..

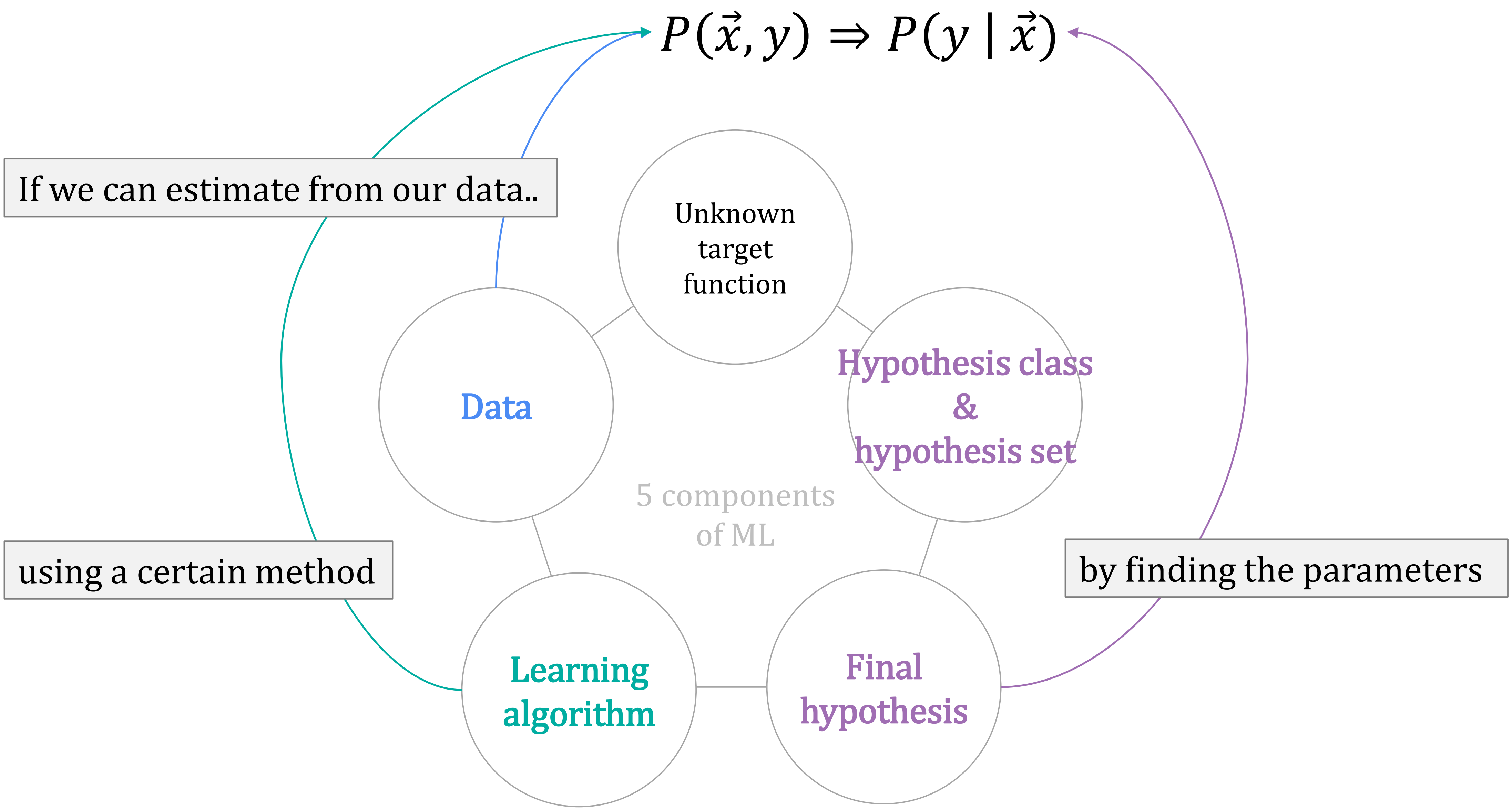


Do you remember?

$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

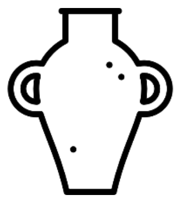


Do you remember?

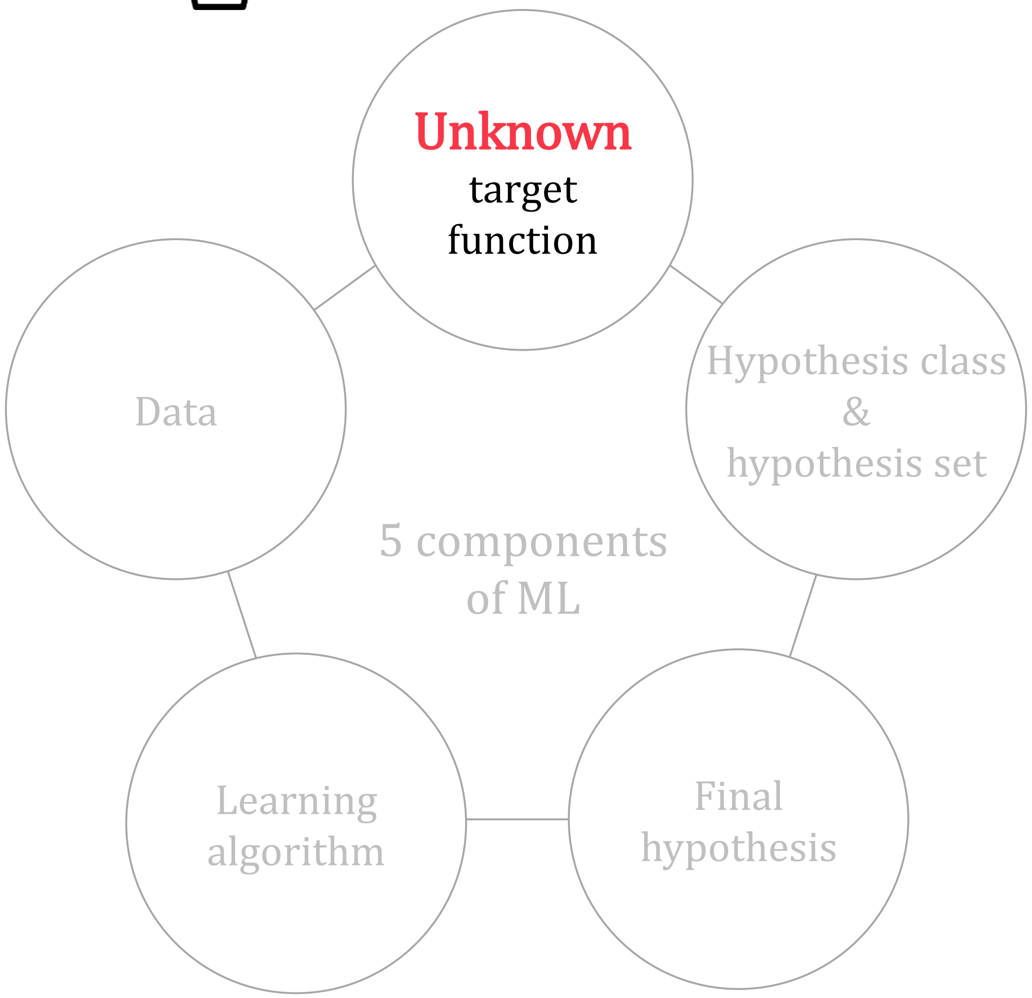


Do you remember?

We don't know
the distribution of
black & white stones,

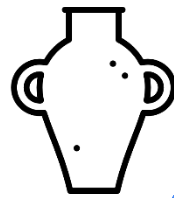


$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$



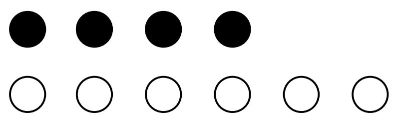
Do you remember?

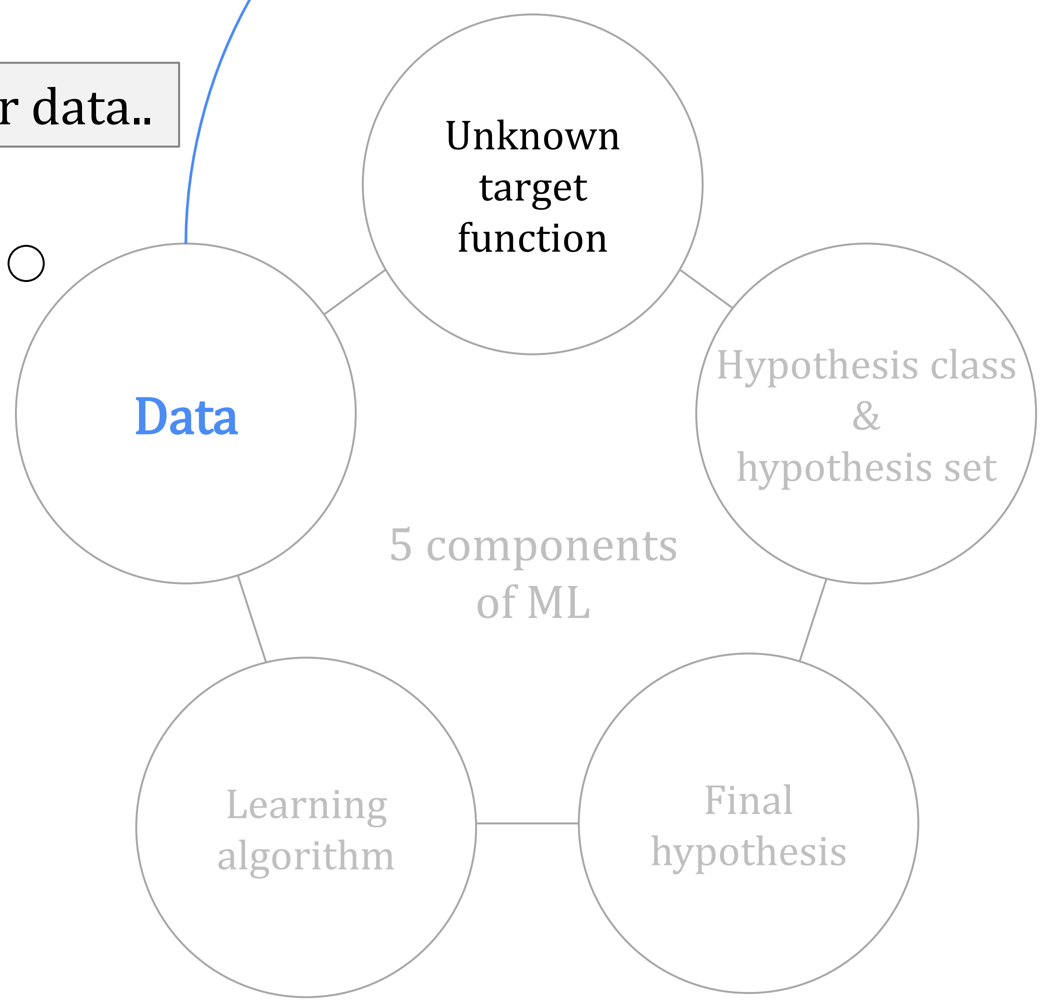
We don't know
the distribution of
black & white stones,



$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

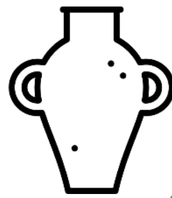
If we can estimate from our data..





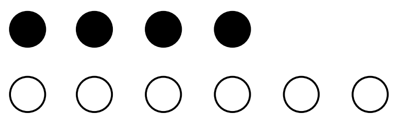
Do you remember?

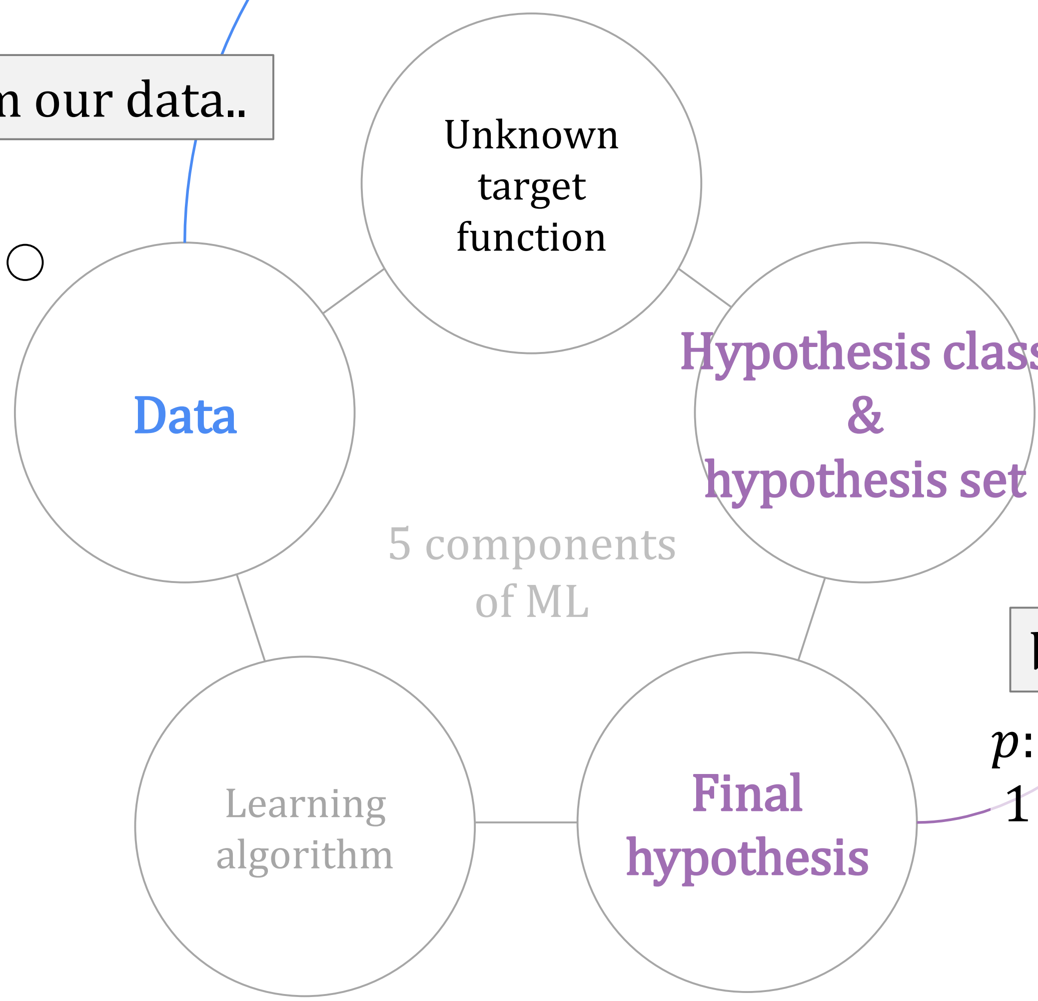
We don't know
the distribution of
black & white stones,



$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

If we can estimate from our data..



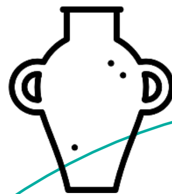


by finding the parameters

p: proportion of black stones
$1 - p$: proportion of white stones

Do you remember?

We don't know
the distribution of
black & white stones,



$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

If we can estimate from our data..

● ● ● ●
○ ○ ○ ○ ○ ○

Unknown
target
function

Data

Hypothesis class
&
hypothesis set

5 components
of ML

Learning
algorithm

Final
hypothesis

using a certain method

$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$

by finding the parameters

p: proportion of black stones
$1-p$: proportion of white stones

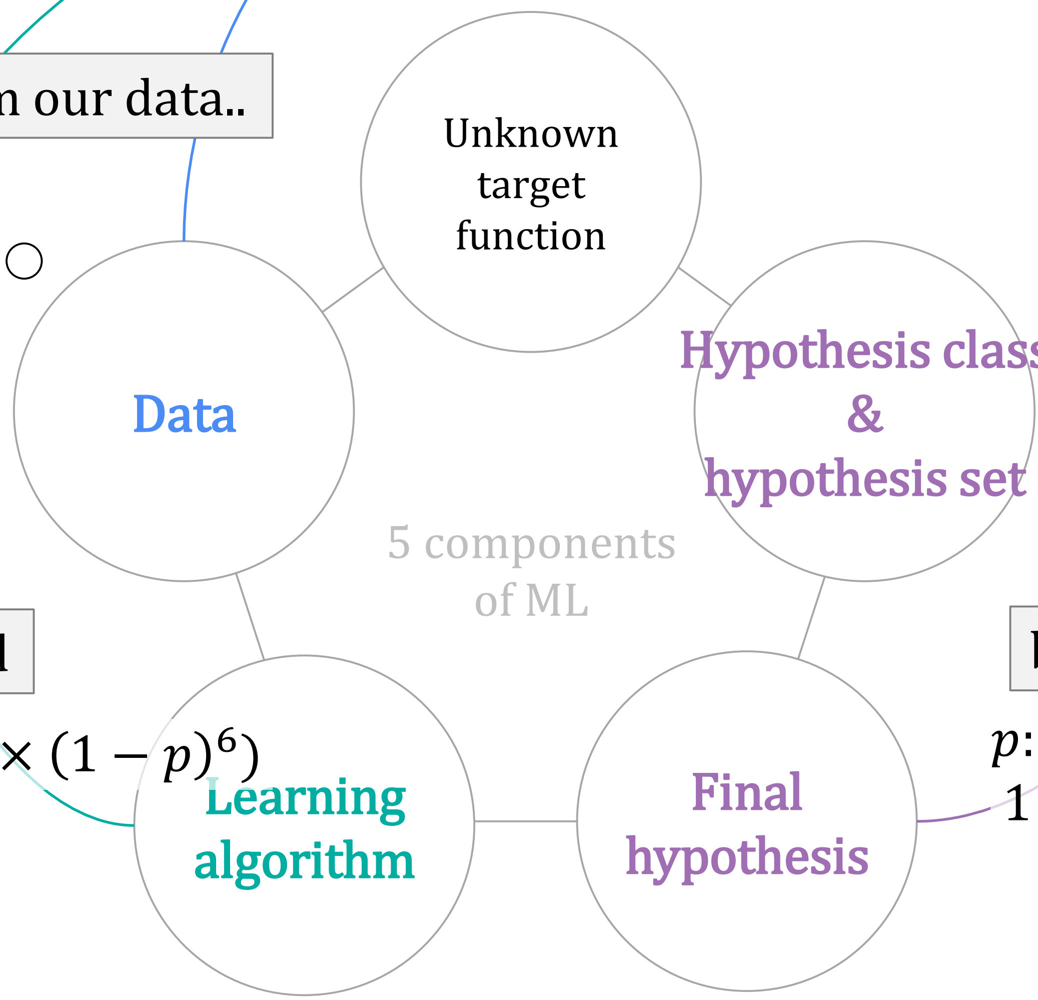
Do you remember?

We don't know
the distribution of
black & white stones,

$$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$$

We can **evaluate the probability of any events**!

If we can estimate from our data..



using a certain method

$$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$$

by finding the parameters

p: proportion of black stones

$1-p$: proportion of white stones

Do you remember?

We don't know the distribution of black & white stones,

$P(\vec{x}, y) \Rightarrow P(y \mid \vec{x})$

We can evaluate the probability of any events!

This is why people say that

MLE is the basic & core **technique** of

the field of **pattern recognition** including

Deep learning, Support Vector Machine, Decision Trees,

Markov Random Field, Neural Networks, Linear Regression,

Logistic Regression, Maximum Entropy Model and etc.

$p^* = \arg max_p \ln(210 \times p^4 \times (1-p)^6)$

p: proportion of black stones

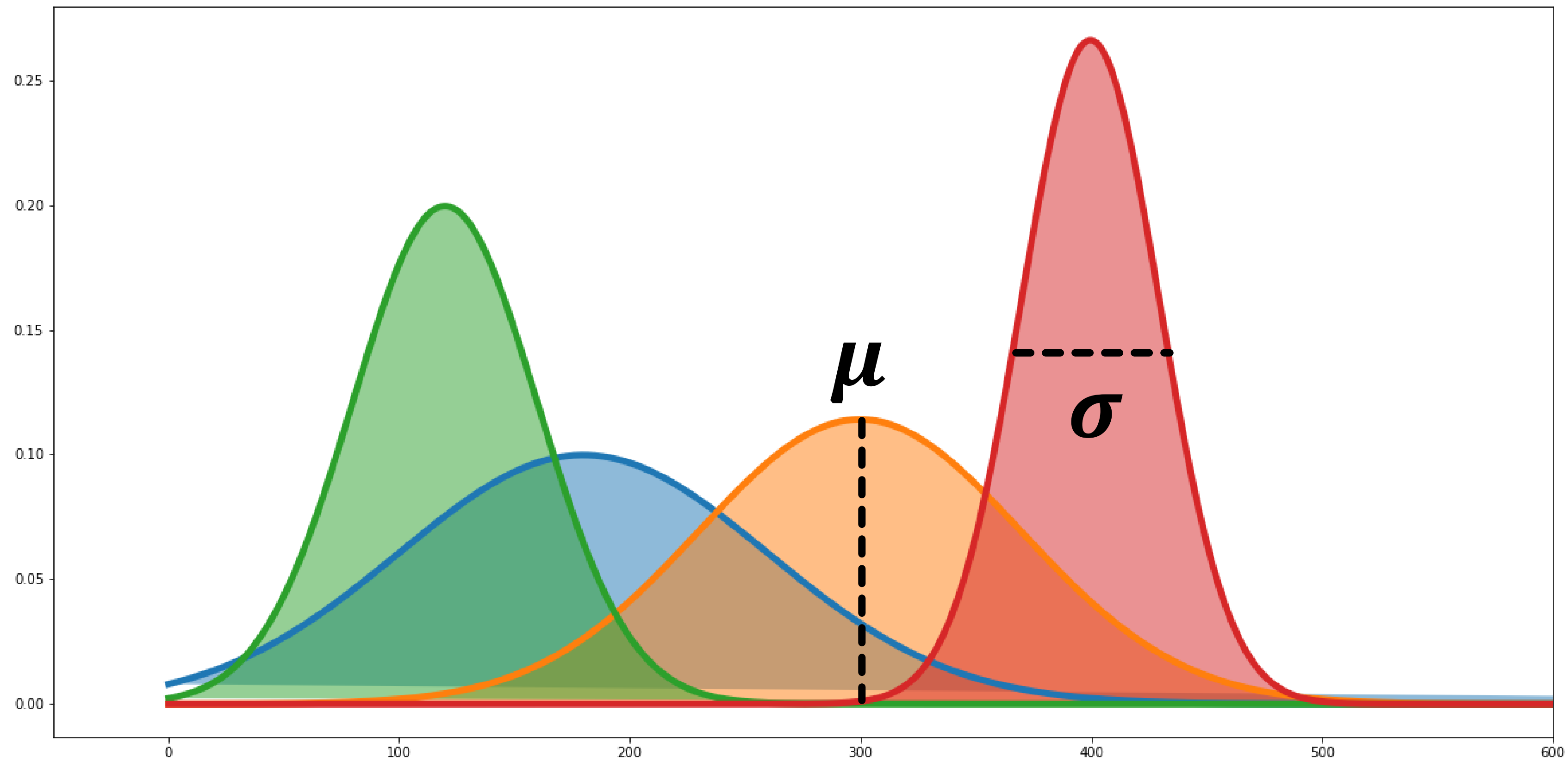
$1 - p$: proportion of white stones

Learning algorithm

Final hypothesis

MLE example #1
Normal distribution

- If we choose normal distribution as our model
  - $P(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
    - Parameters: μ(mean), σ(standard deviation)



MLE example #1

Normal distribution

Why μ and σ?

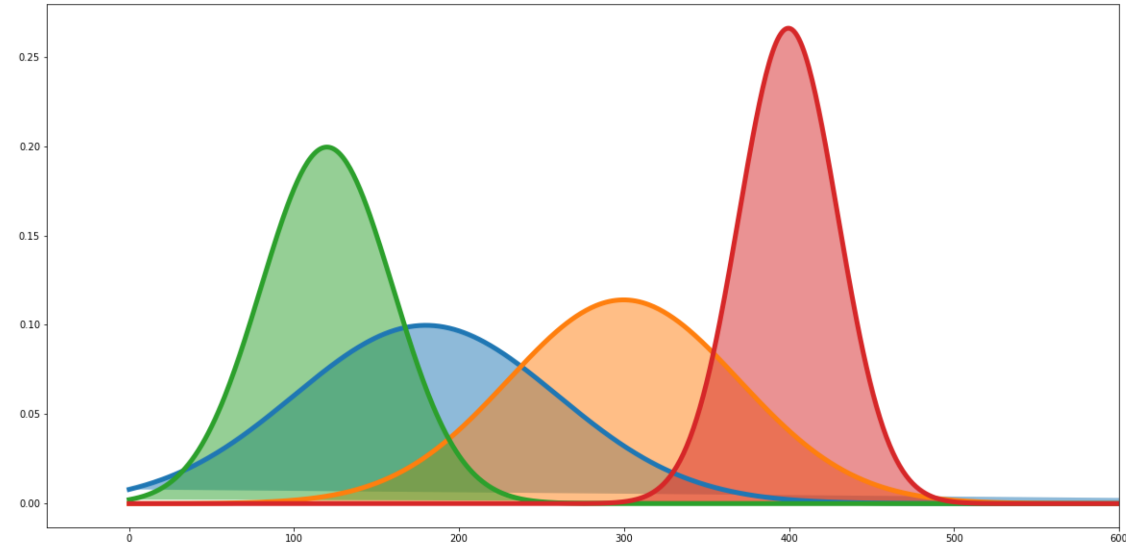
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Why not the other values ?

This is just a **model**!

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

We can get **diverse normal curves** by plugging in various values into α and β



MLE example #1

Normal distribution

Model

$$\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

MLE example #1

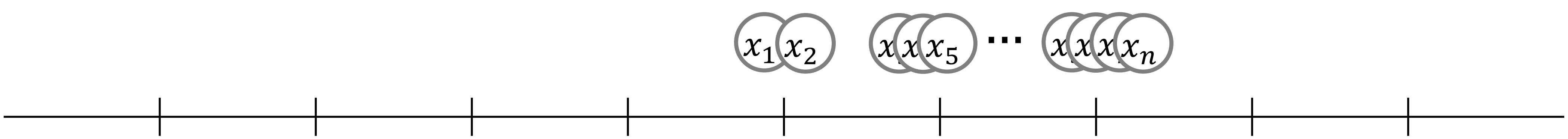
Normal distribution

Model

$$\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

Data

$$x_1, x_2, x_3, \dots, x_n$$



MLE example #1

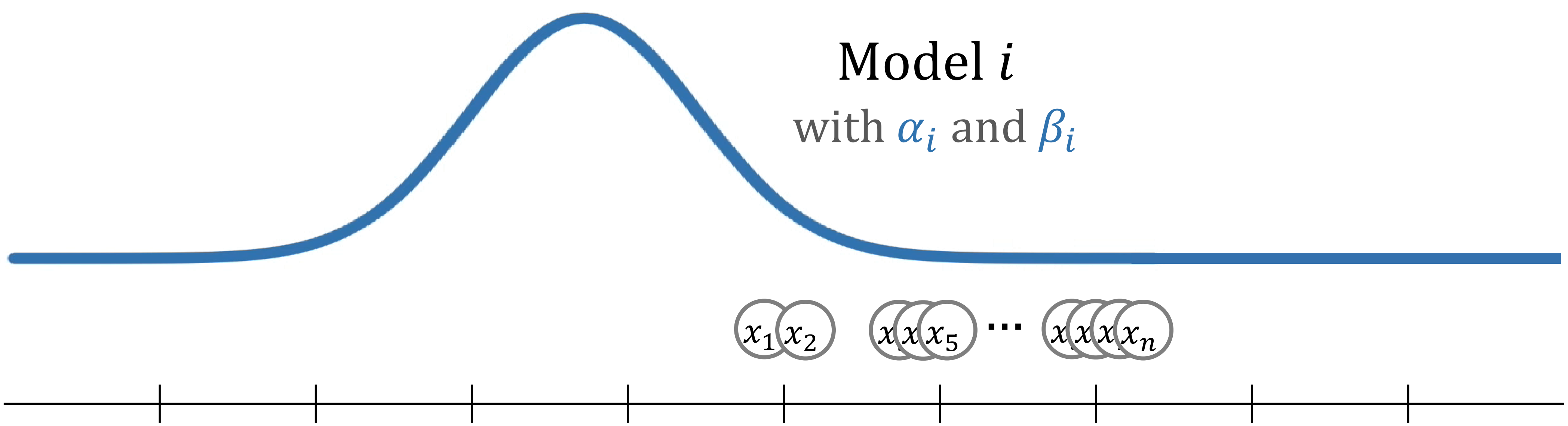
Normal distribution

Model

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

Data

$$x_1, x_2, x_3, \dots, x_n$$



MLE example #1

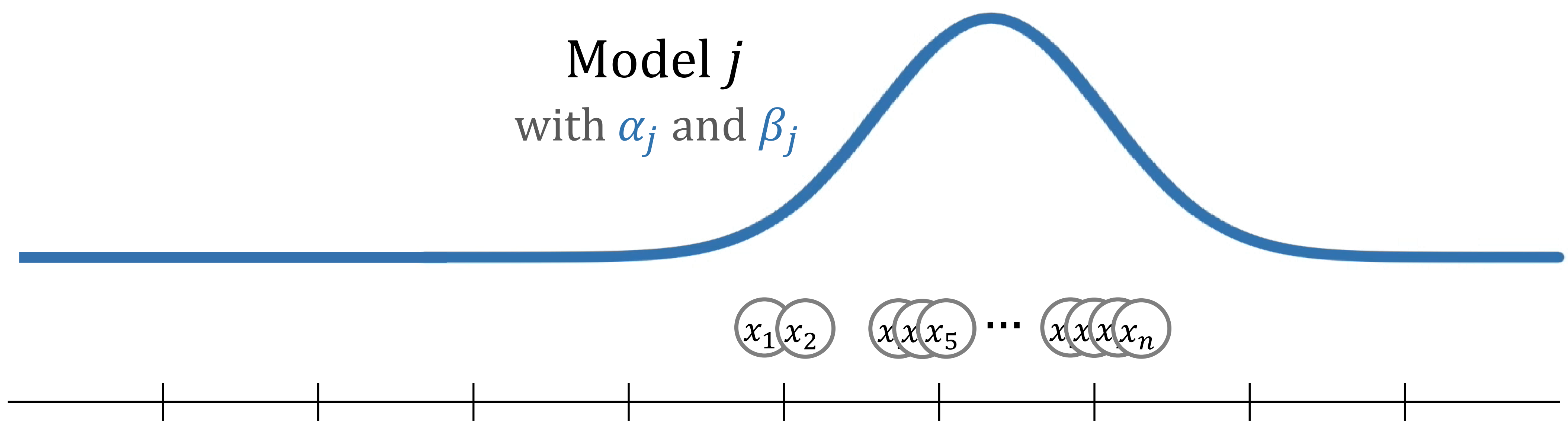
Normal distribution

Model

$$\frac{1}{\sqrt{2\pi\alpha^2}}\, e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

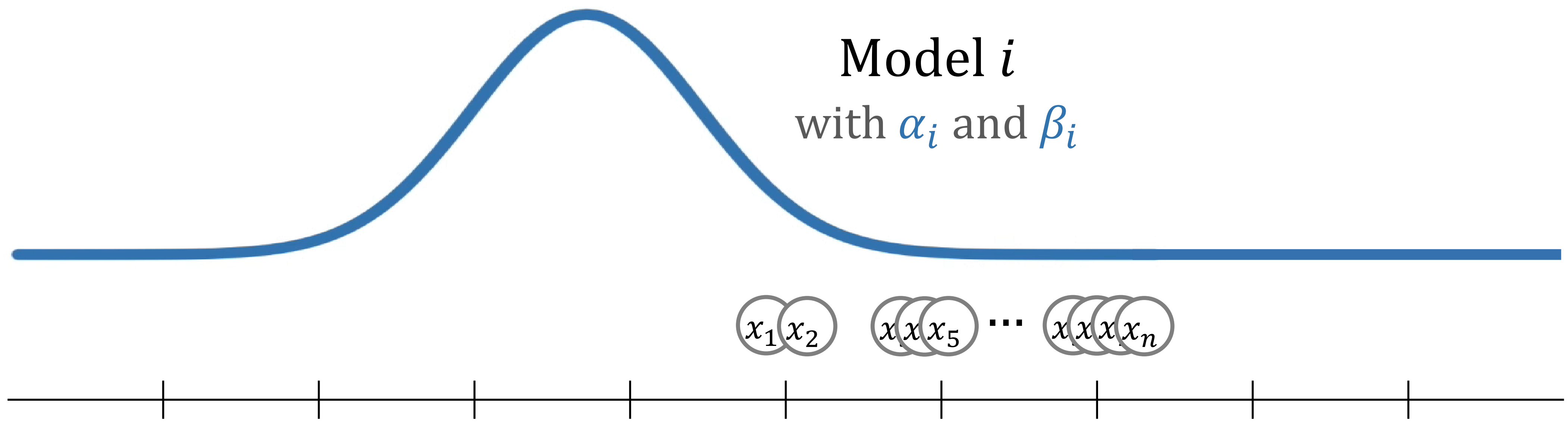
Data

$$x_1, x_2, x_3, \dots, x_n$$

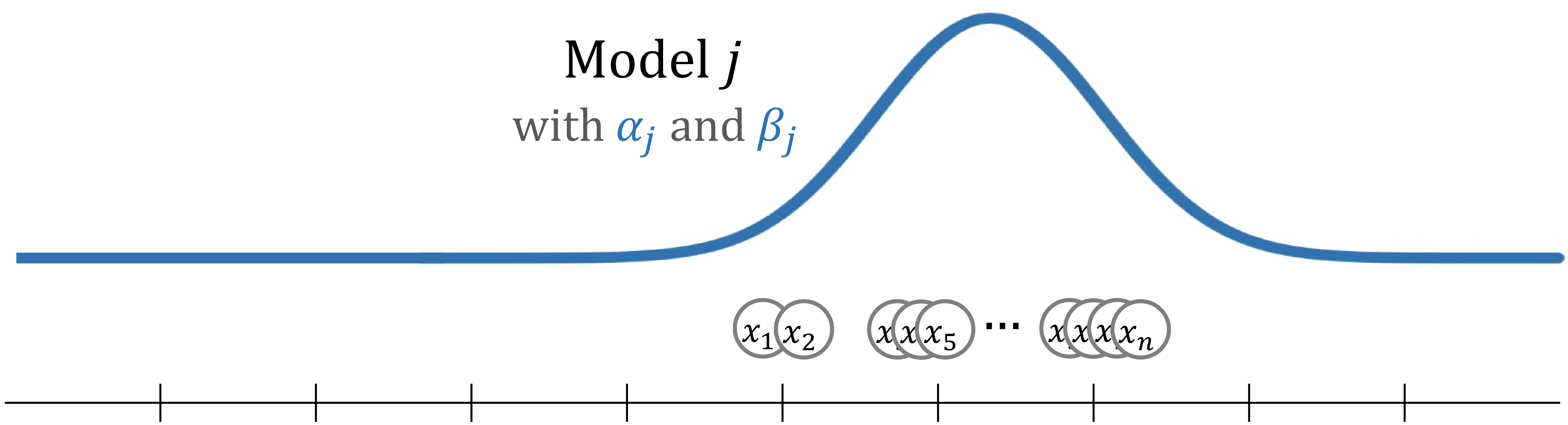


MLE example #1

Normal distribution

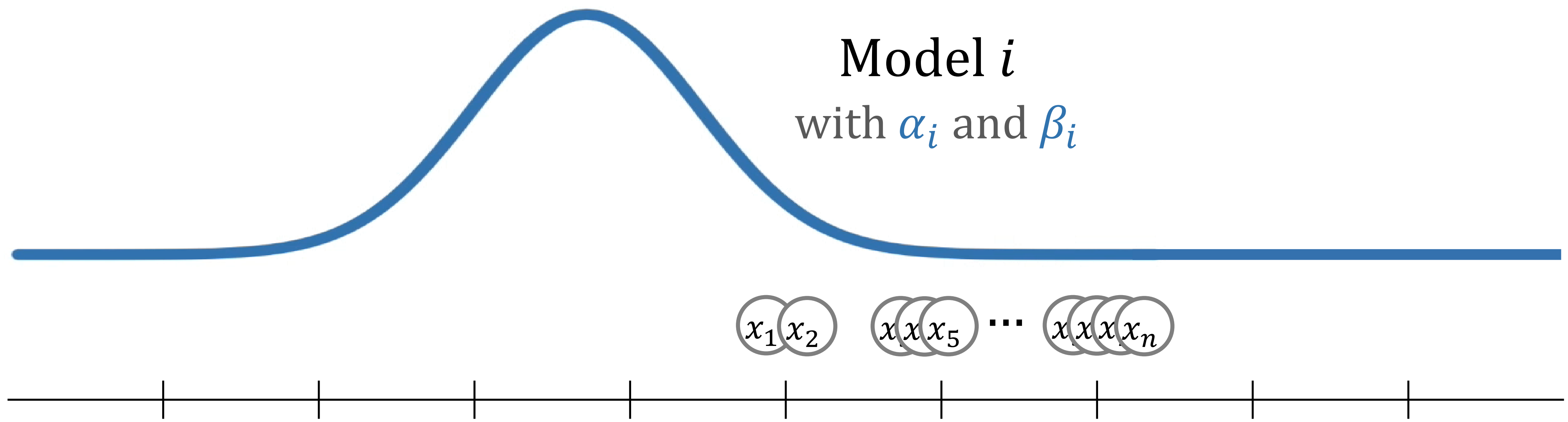


Which model (α & β) is better and why?

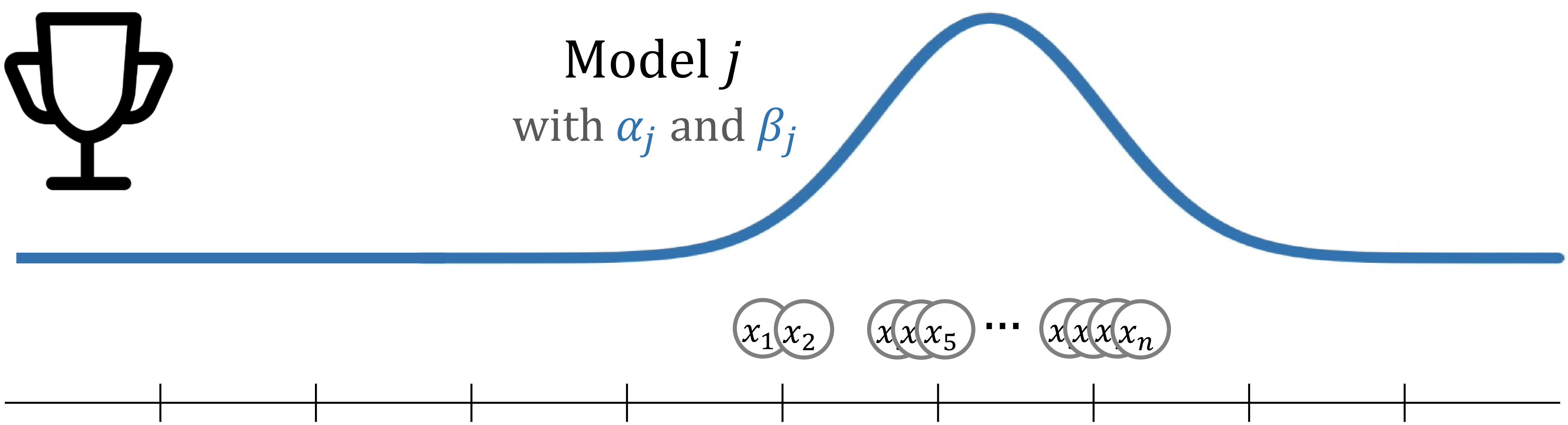


MLE example #1

Normal distribution

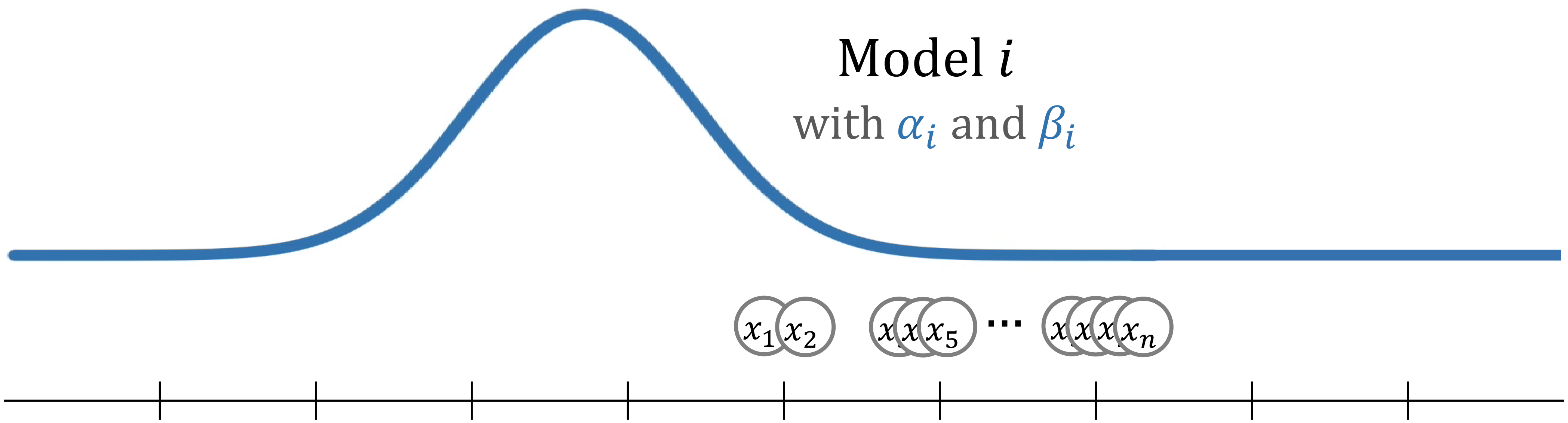


Our observation is common → High probability

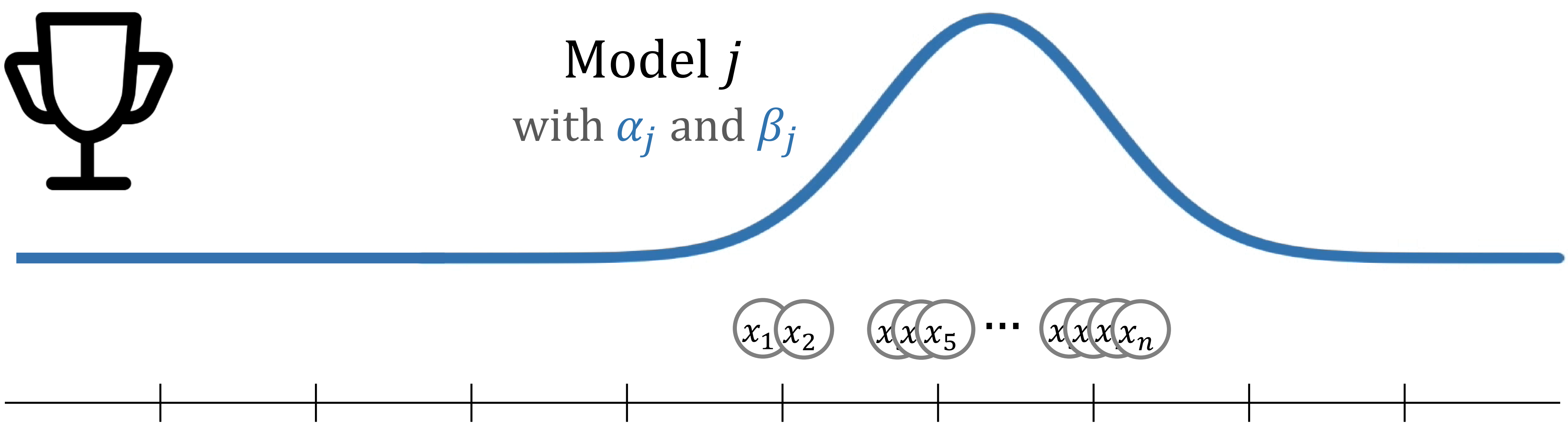


MLE example #1

Normal distribution



$$L(\alpha_i, \beta_i \mid x_1, x_2, x_3, \dots, x_n) < L\left(\alpha_j, \beta_j \mid x_1, x_2, x_3, \dots, x_n\right)$$



MLE example #1

Normal distribution

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

$$L(\alpha, \beta \mid x_1, x_2, \dots, x_n)$$

How to evaluate the likelihood ?

MLE example #1

Normal distribution

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

$$L(\alpha, \beta \mid x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_2-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

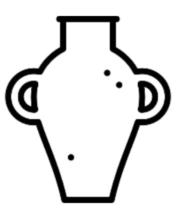
MLE example #1

Normal distribution

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

$$L(\alpha, \beta \mid x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_2-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

p ● ○ … ● $= p \times (1-p) \times \ \dots \times p$



The event, sampling each stone, is **independent** to each other.

MLE example #1

Normal distribution

$$L(\alpha, \beta \mid x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_2-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

$$\alpha^*, \beta^* = \arg max_{\alpha,\beta} \left(\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2} \right)$$

MLE example #1

Normal distribution

Time to find α and β that maximize

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

MLE example #1

Normal distribution

Time to find α and β that maximize

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

using partial derivative w.r.t α and β

MLE example #1

Normal distribution

Time to find α and β that maximize

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}$$

using partial derivative w.r.t α and β & natural logarithm (log_e = ln)

MLE example #1

Normal distribution

Time to find α and β that maximize

$$\ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

using partial derivative w.r.t α and β

MLE example #1

Normal distribution

$$\ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

$$= \ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2}\right) + \cdots + \ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

MLE example #1

Normal distribution

$$= \ln\left(\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2}\right) + \cdots + \ln\left(\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

$$\ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}\right) + \ln\left(e^{-\frac{1}{2}\times\frac{(x_1-\beta)^2}{\alpha^2}}\right)$$

$$= \ln\left(\frac{1}{(2\pi\alpha^2)^{\frac{1}{2}}}\right) + \ln\left(e^{-\frac{(x_1-\beta)^2}{2\alpha^2}}\right)$$

$$= \ln\left((2\pi\alpha^2)^{-\frac{1}{2}}\right) - \frac{(x_1-\beta)^2}{2\alpha^2}\ln(e)$$

$$= -\frac{1}{2}\ln(2\pi\alpha^2) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

$$= -\frac{1}{2}\left(\ln(2\pi) + \ln(\alpha^2)\right) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\alpha^2) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

$$= -\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

MLE example #1

Normal distribution

$$\ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2}\right)+\cdots+\ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

$$=-\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}+\cdots+-\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_n-\beta)^2}{2\alpha^2}$$

MLE example #1

Normal distribution

$$ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_1-\beta}{\alpha}\right)^2}\right)+\cdots+ln\left(\frac{1}{\sqrt{2\pi\alpha^2}}e^{-\frac{1}{2}\left(\frac{x_n-\beta}{\alpha}\right)^2}\right)$$

$$= -\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}+\cdots+-\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_n-\beta)^2}{2\alpha^2}$$

$$\begin{aligned}= & -\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}\\ & -\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_2-\beta)^2}{2\alpha^2}\\ & \ldots\\ & -\frac{1}{2}\ln(2\pi)-\ln(\alpha)-\frac{(x_n-\beta)^2}{2\alpha^2}\end{aligned}$$

MLE example #1

Normal distribution

$$
\begin{aligned}
= & \boxed{\begin{array}{l} -\frac{1}{2}\ln(2\pi) \\ -\frac{1}{2}\ln(2\pi) \\ \ldots \\ -\frac{1}{2}\ln(2\pi) \end{array}} \begin{array}{l} -\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2} \\ -\ln(\alpha)-\frac{(x_2-\beta)^2}{2\alpha^2} \\ \\ -\ln(\alpha)-\frac{(x_n-\beta)^2}{2\alpha^2} \end{array}
\end{aligned}
$$

$$
\downarrow
$$

$$
-\frac{n}{2}\ln(2\pi)
$$

MLE example #1

Normal distribution

$$= -\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

$$-\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_2-\beta)^2}{2\alpha^2}$$

$$\dots$$

$$-\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_n-\beta)^2}{2\alpha^2}$$

$$-\frac{n}{2}\ln(2\pi) - n\ln(\alpha)$$

MLE example #1

Normal distribution

$$= -\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2}$$

$$-\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_2-\beta)^2}{2\alpha^2}$$

$$\dots$$

$$-\frac{1}{2}\ln(2\pi) - \ln(\alpha) - \frac{(x_n-\beta)^2}{2\alpha^2}$$

$$-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \dots - \frac{(x_n-\beta)^2}{2\alpha^2}$$

MLE example #1

Normal distribution

Likelihood function

$$-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \cdots - \frac{(x_n-\beta)^2}{2\alpha^2}$$

As it has two variables α and β,

$$\frac{\partial}{\partial\alpha}\left(-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \cdots - \frac{(x_n-\beta)^2}{2\alpha^2}\right)$$

&

$$\frac{\partial}{\partial\beta}\left(-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \cdots - \frac{(x_n-\beta)^2}{2\alpha^2}\right)$$

MLE example #1
Normal distribution

- Partial derivative w.r.t β

$$\boxed{\frac{\partial}{\partial\beta}\left(-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \cdots - \frac{(x_n-\beta)^2}{2\alpha^2}\right)}$$

$$= \frac{\partial}{\partial\beta}\left(-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{1}{2\alpha^2}({x_1}^2 - 2x_1\beta + \beta^2) - \cdots - \frac{1}{2\alpha^2}({x_n}^2 - 2x_n\beta + \beta^2)\right)$$

$$= 0 - 0 - \frac{1}{2\alpha^2}(-2x_1 + 2\beta) - \cdots - \frac{1}{2\alpha^2}(-2x_n + 2\beta)$$

$$= \frac{1}{\alpha^2}(x_1 - \beta) + \cdots + \frac{1}{\alpha^2}(x_n - \beta)$$

$$= \frac{(x_1 - \beta) + \cdots + (x_n - \beta)}{\alpha^2}$$

MLE example #1

Normal distribution

- Partial derivative w.r.t β

$$\boxed{\frac{\partial}{\partial\beta}\left(-\frac{n}{2}\ln(2\pi) - n\ln(\alpha) - \frac{(x_1-\beta)^2}{2\alpha^2} - \cdots - \frac{(x_n-\beta)^2}{2\alpha^2}\right)}$$

$$= \frac{(x_1-\beta) + \cdots + (x_n-\beta)}{\alpha^2}$$

What we are interested in→ $$\frac{(x_1-\beta) + \cdots + (x_n-\beta)}{\alpha^2} = 0$$

$$(x_1-\beta) + \cdots + (x_n-\beta) = 0$$

$$x_1 + x_2 + \cdots + x_n - n\beta = 0$$

$$x_1 + x_2 + \cdots + x_n = n\beta$$

We call it 'mean'
& denote as μ → $$\frac{x_1 + x_2 + \cdots + x_n}{n} = \beta$$

MLE example #1

Normal distribution

- Partial derivative w.r.t α

$$\boxed{\frac{\partial}{\partial\alpha}\left(-\frac{n}{2}\ln(2\pi)-n\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}-\cdots-\frac{(x_n-\beta)^2}{2\alpha^2}\right)}$$

$$=\frac{\partial}{\partial\alpha}\left(-\frac{n}{2}\ln(2\pi)-n\ln(\alpha)-\frac{(x_1-\beta)^2}{2}\alpha^{-2}-\cdots-\frac{(x_n-\beta)^2}{2}\alpha^{-2}\right)$$

$$=0-\frac{n}{\alpha}-\frac{(x_1-\beta)^2}{2}(-2)\alpha^{-3}-\cdots-\frac{(x_n-\beta)^2}{2}(-2)\alpha^{-3}$$

$$=-\frac{n}{\alpha}+(x_1-\beta)^2\alpha^{-3}+\cdots+(x_n-\beta)^2\alpha^{-3}$$

$$=-\frac{n}{\alpha}+\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{\alpha^3}$$

MLE example #1

Normal distribution

- Partial derivative w.r.t α

$$\boxed{\frac{\partial}{\partial\alpha}\left(-\frac{n}{2}\ln(2\pi)-n\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}-\cdots-\frac{(x_n-\beta)^2}{2\alpha^2}\right)}$$

$$=-\frac{n}{\alpha}+\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{\alpha^3}$$

What we are interested in→ $-\frac{n}{\alpha}+\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{\alpha^3}=0$

$$\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{\alpha^3}=\frac{n}{\alpha}$$

$$\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{\alpha^2}=n$$

$$(x_1-\beta)^2+\cdots+(x_n-\beta)^2=\alpha^2 n$$

MLE example #1

Normal distribution

- Partial derivative w.r.t α

$$\boxed{\frac{\partial}{\partial\alpha}\left(-\frac{n}{2}\ln(2\pi)-n\ln(\alpha)-\frac{(x_1-\beta)^2}{2\alpha^2}-\cdots-\frac{(x_n-\beta)^2}{2\alpha^2}\right)}$$

$$(x_1-\beta)^2+\cdots+(x_n-\beta)^2=\alpha^2 n$$

$$\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{n}=\alpha^2$$

We call it 'standard deviation' & denote as σ $\rightarrow$ $\sqrt{\frac{(x_1-\beta)^2+\cdots+(x_n-\beta)^2}{n}}=\alpha$

MLE example #1

Normal distribution

Model

Normal distribution

$$\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2}\left(\frac{x-\beta}{\alpha}\right)^2}$$

Data

$$x_1, x_2, x_3, \dots, x_n$$

Likelihood of any given data is maximal under the normal distribution model where $\alpha = \sigma$ and $\beta = \mu$

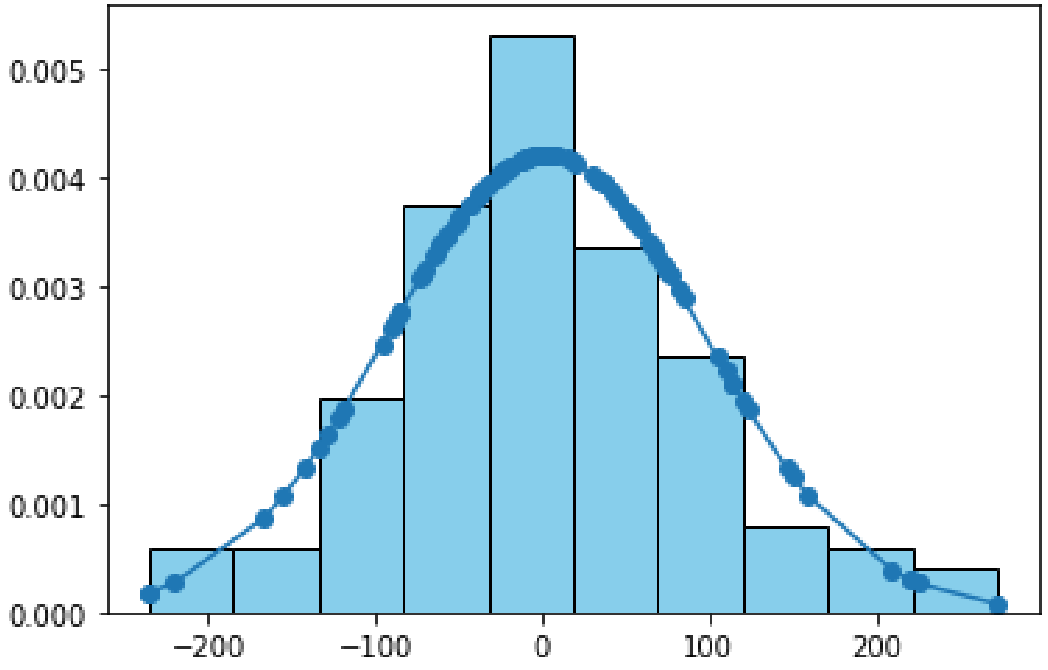
and therefore,

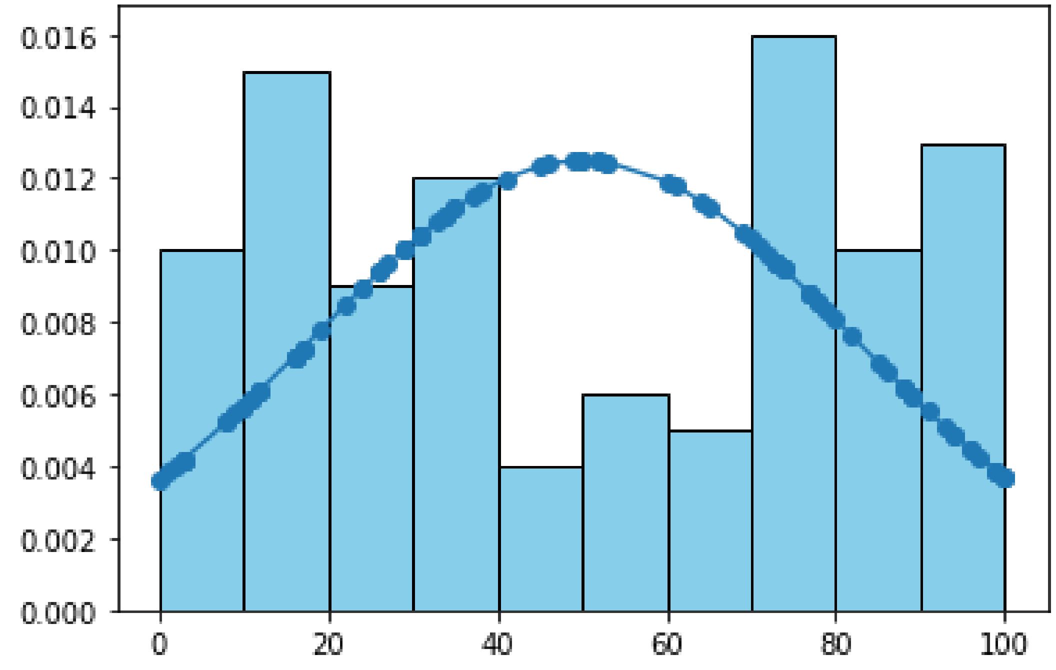
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

MLE example #1

Normal distribution

Fitted normal distribution model for each dataset





Extra quiz

- Have you ever thought about how numpy.random.normal works?

numpy.random.**normal**(*loc=0.0*, *scale=1.0*, *size=None*)

Draw random samples from a normal (Gaussian) distribution.

The probability density function of the normal distribution, first derived by De Moivre and 200 years later by both Gauss and Laplace independently [2], is often called the bell curve because of its characteristic shape (see the example below).

The normal distributions occurs often in nature. For example, it describes the commonly occurring distribution of samples influenced by a large number of tiny, random disturbances, each with its own unique distribution [2].

> **Note:**
> New code should use the `normal` method of a `default_rng()` instance instead; see *random-quick-start*.

Parameters:

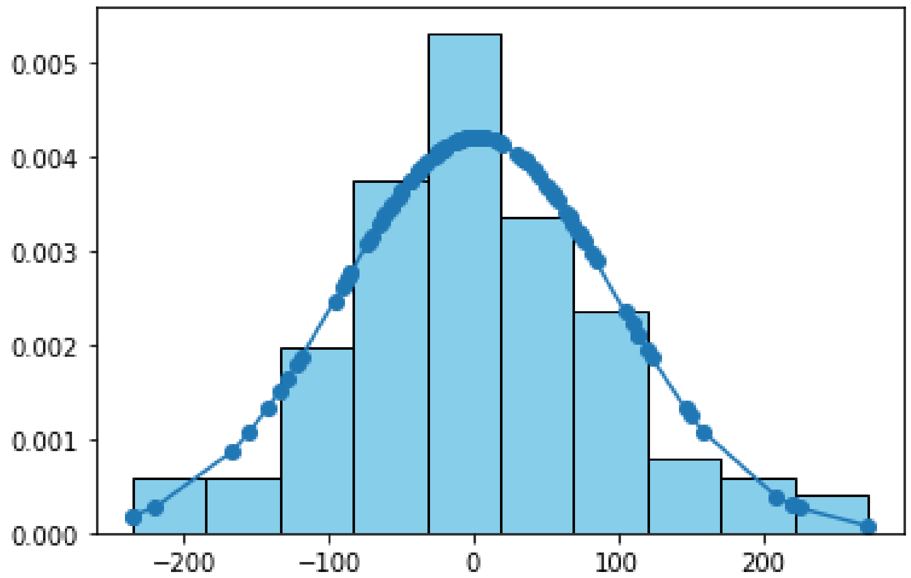
loc : *float or array_like of floats*

Mean ("centre") of the distribution.

scale : *float or array_like of floats*

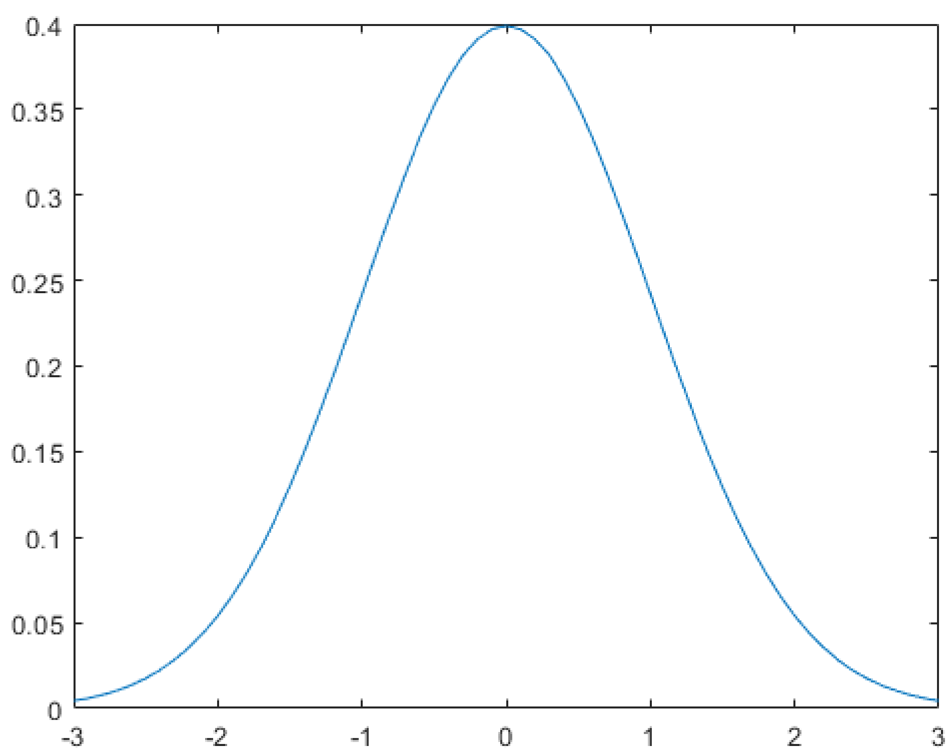
Standard deviation (spread or "width") of the distribution. Must be non-negative.

size : *int or tuple of ints, optional*

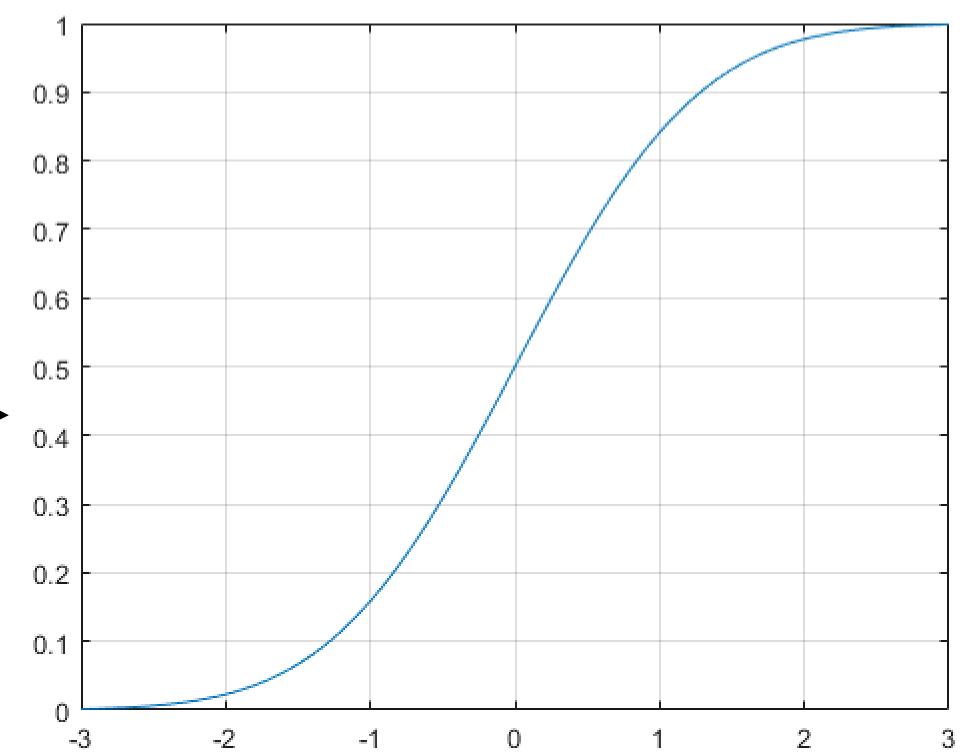


Extra quiz

- Have you ever thought about how numpy.random.normal works?



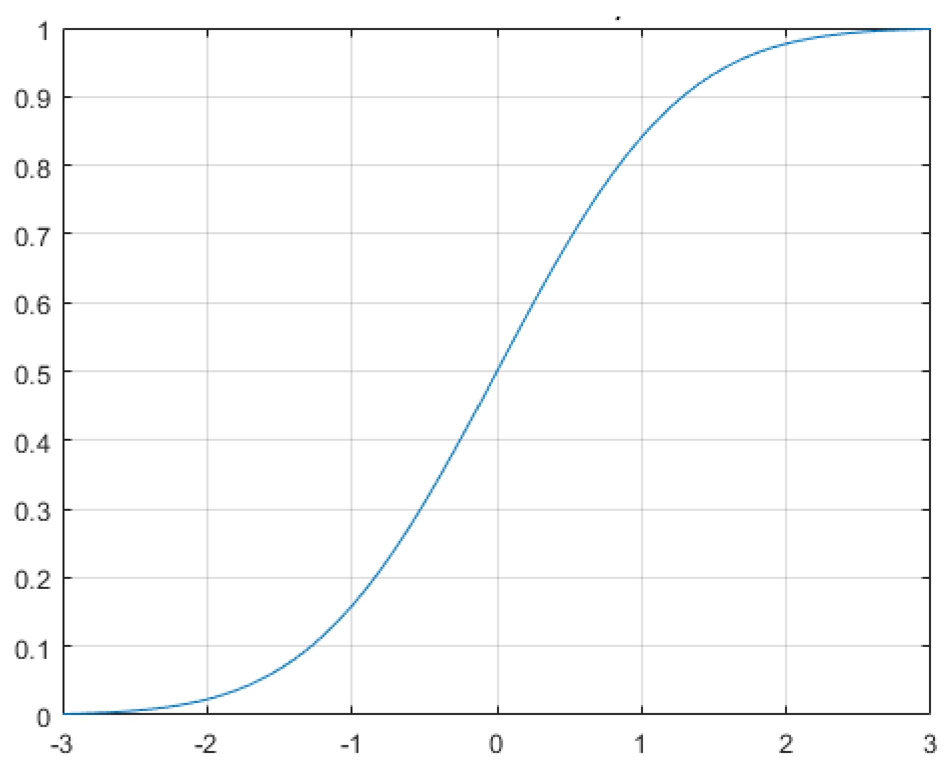
Normal distribution



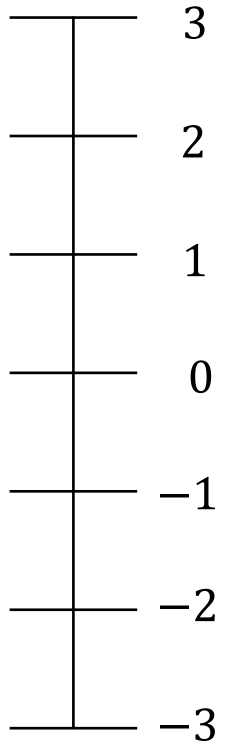
Cumulative normal distribution

Extra quiz

- Have you ever thought about how numpy.random.normal works?



Cumulative normal distribution

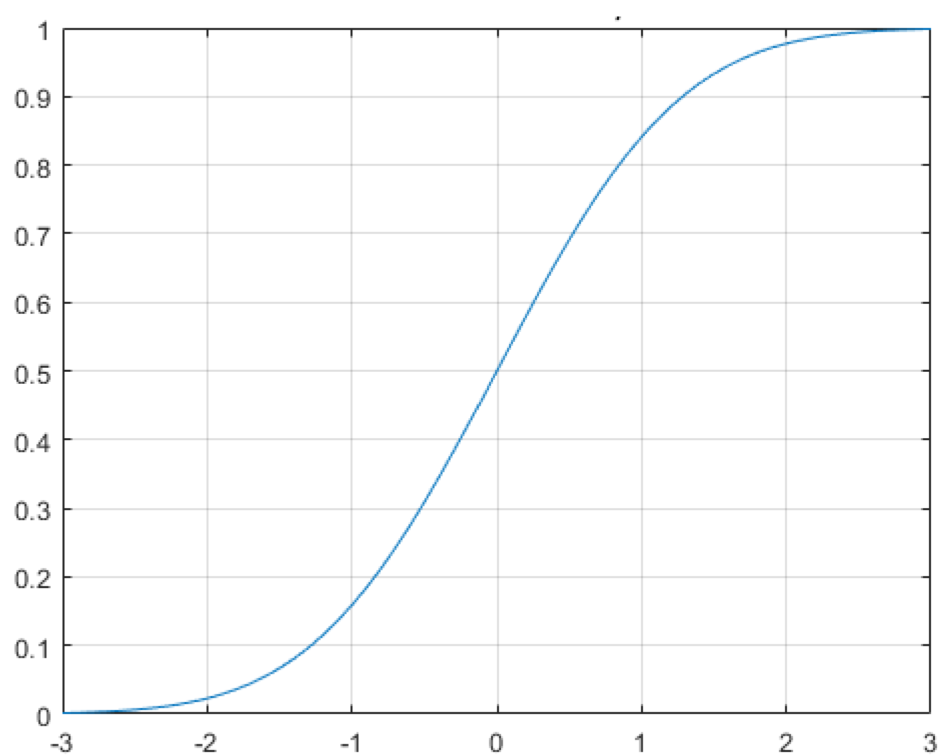


numpy.random

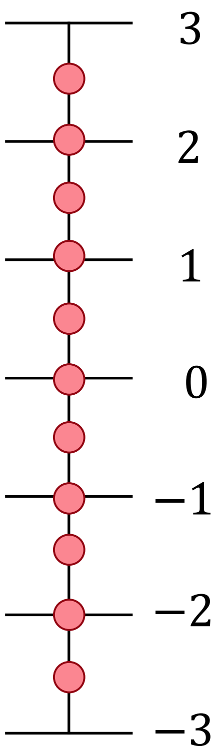
Uniform distribution

Extra quiz

- Have you ever thought about how numpy.random.normal works?



Cumulative normal distribution

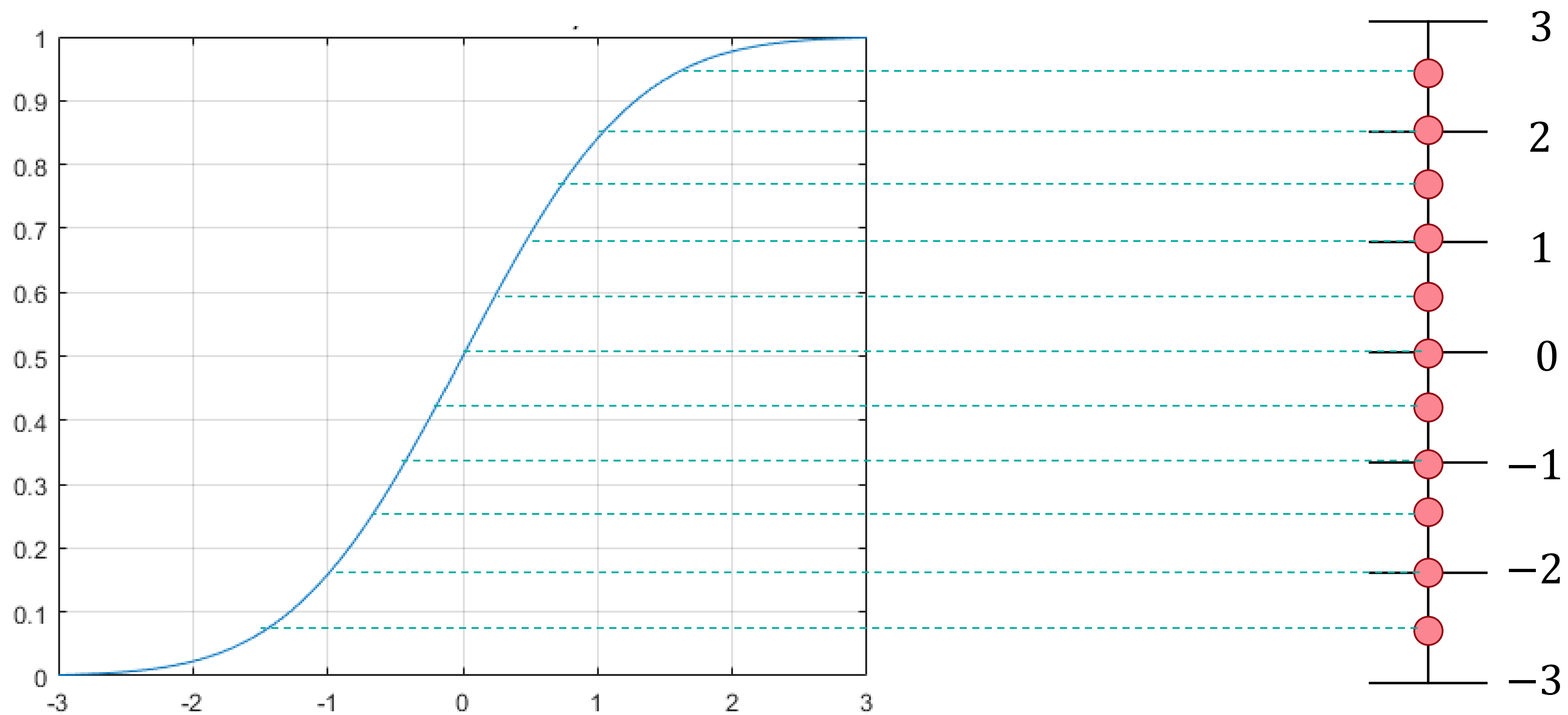


numpy.random

Uniform distribution

Extra quiz

- Have you ever thought about how numpy.random.normal works?



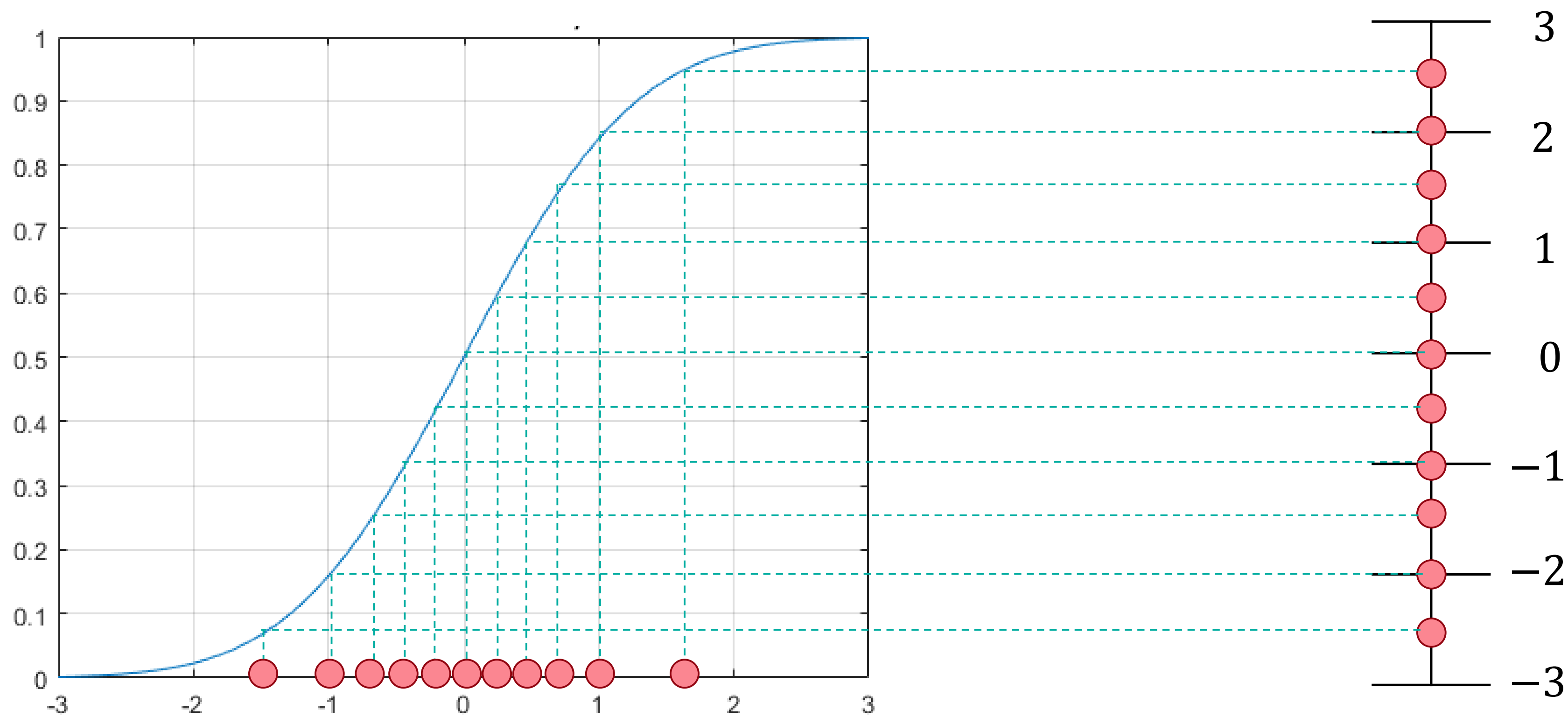
Cumulative normal distribution

numpy.random

Uniform distribution

Extra quiz

- Have you ever thought about how numpy.random.normal works?



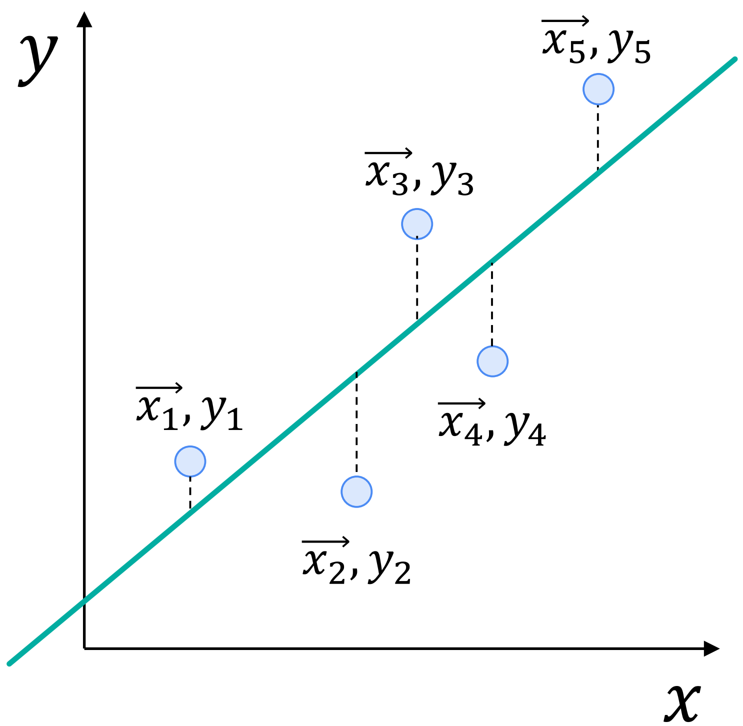
Cumulative normal distribution

numpy.random

Uniform distribution

MLE example #2

Linear regression



$\boxed{y = \alpha x + \beta}$

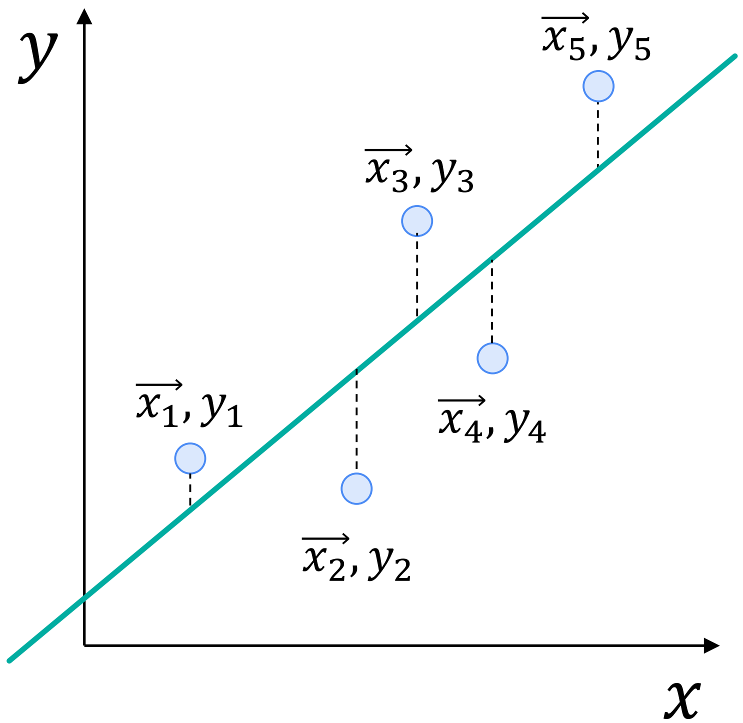
We need to find α and β minimizing

$$\sum_{i=1}^{5} (y_i - (\alpha x_i + \beta))^2$$

Let's call it 'ML approach'.

MLE example #2

Linear regression



$y = \alpha x + \beta$

We need to find α and β minimizing

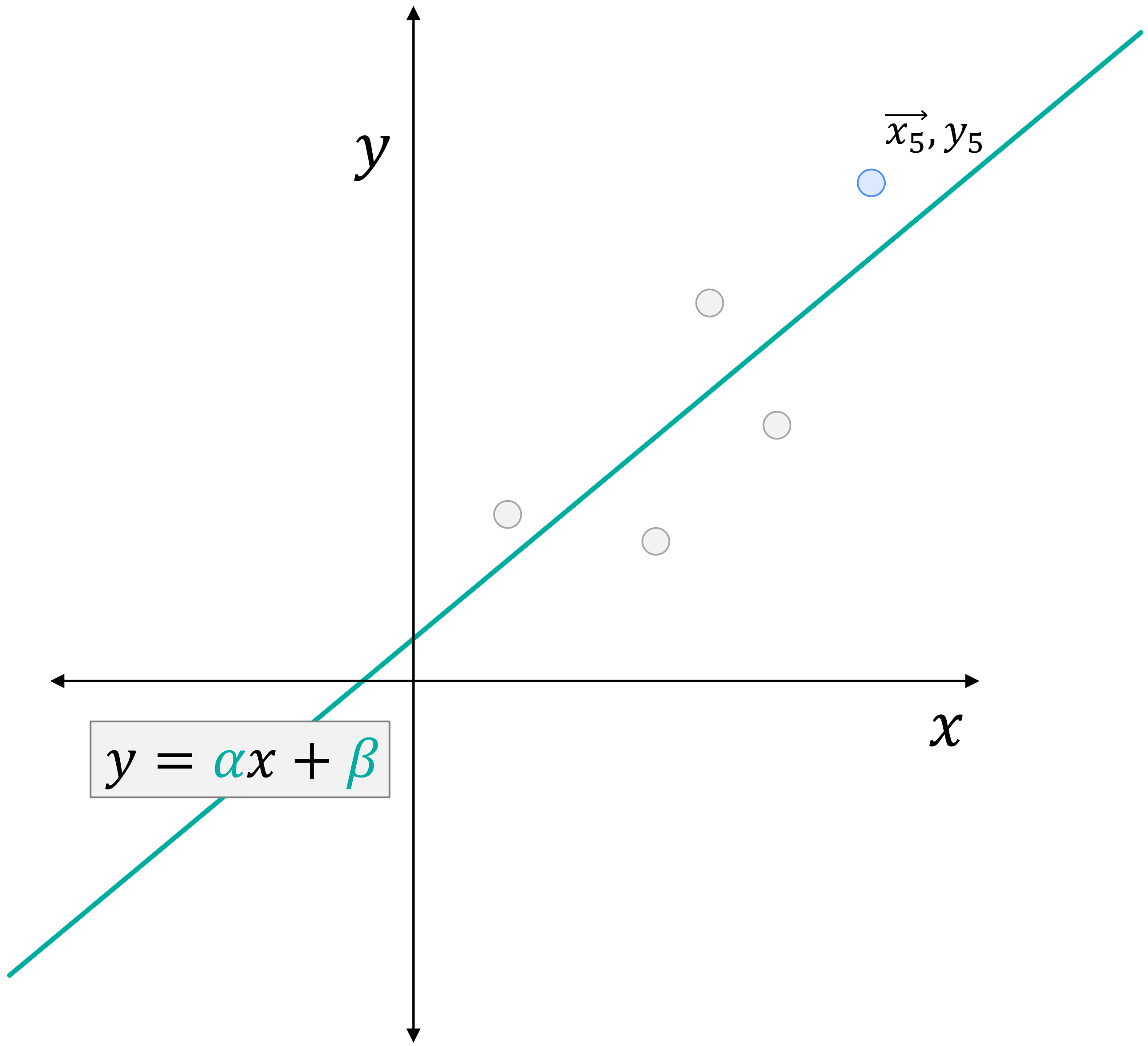
$$\sum_{i=1}^{5} (y_i - (\alpha x_i + \beta))^2$$

Let's call it 'ML approach'.

Let's do w.r.t. 'MLE'.

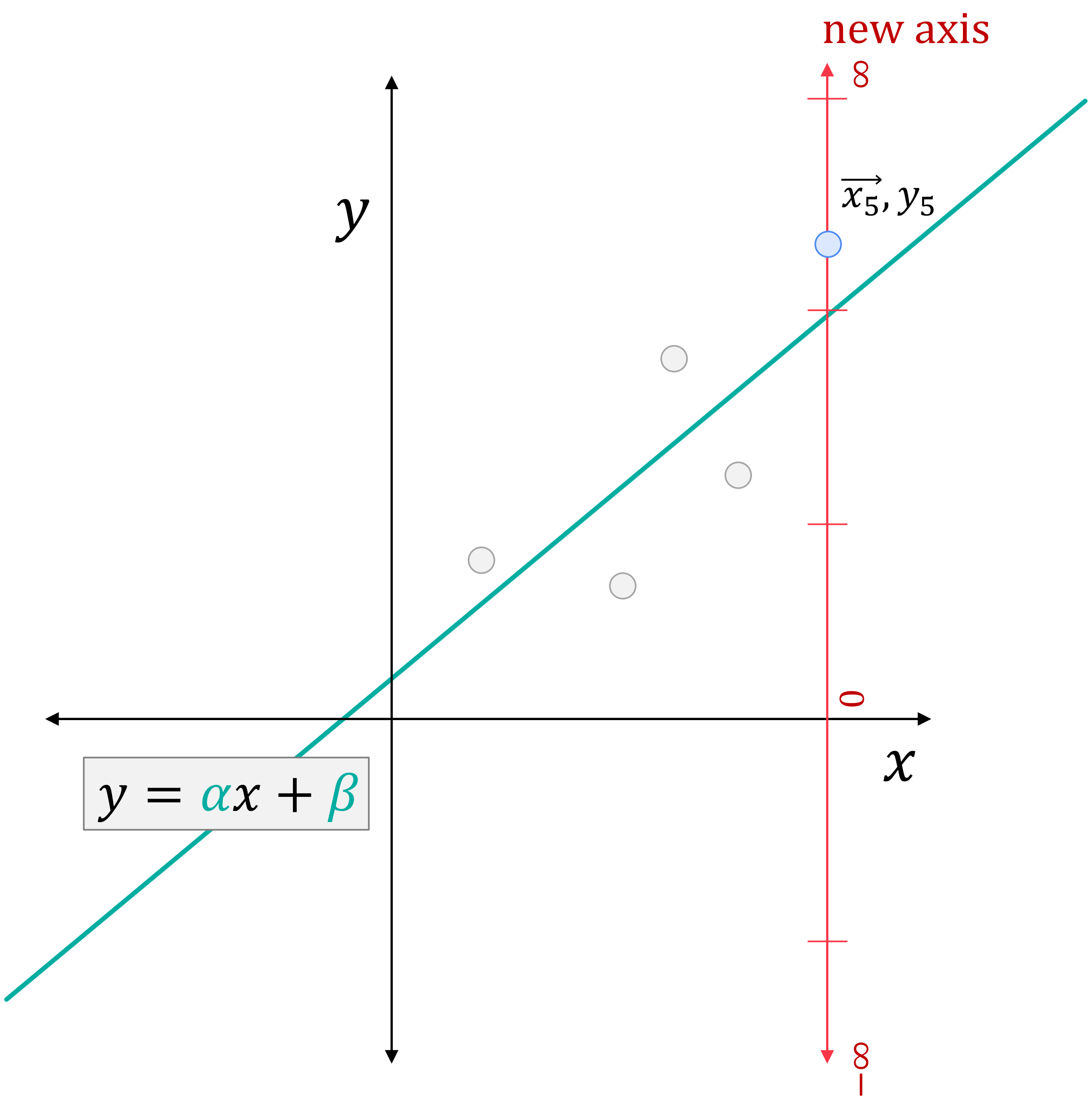
MLE example #2

Linear regression



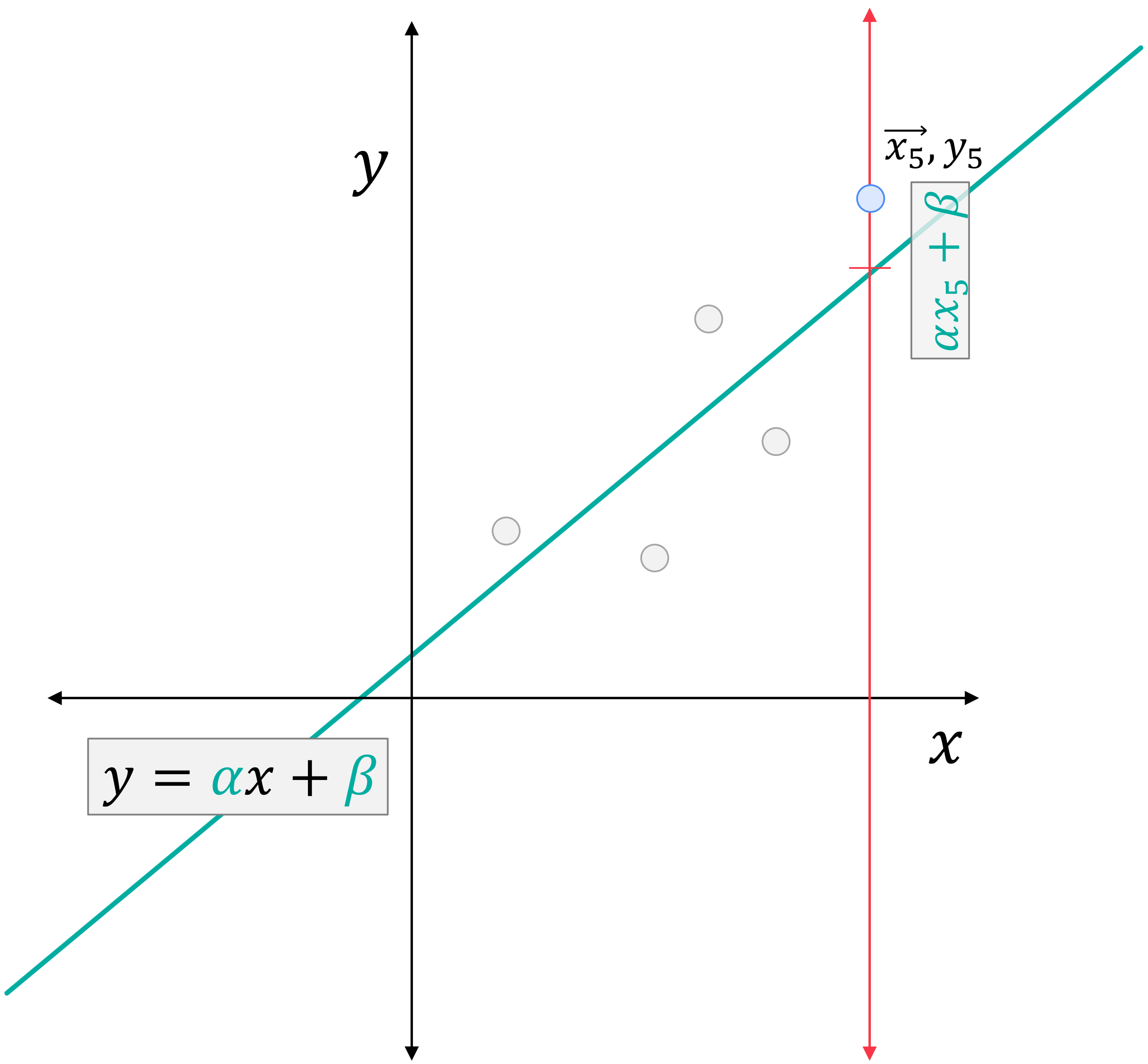
MLE example #2

Linear regression



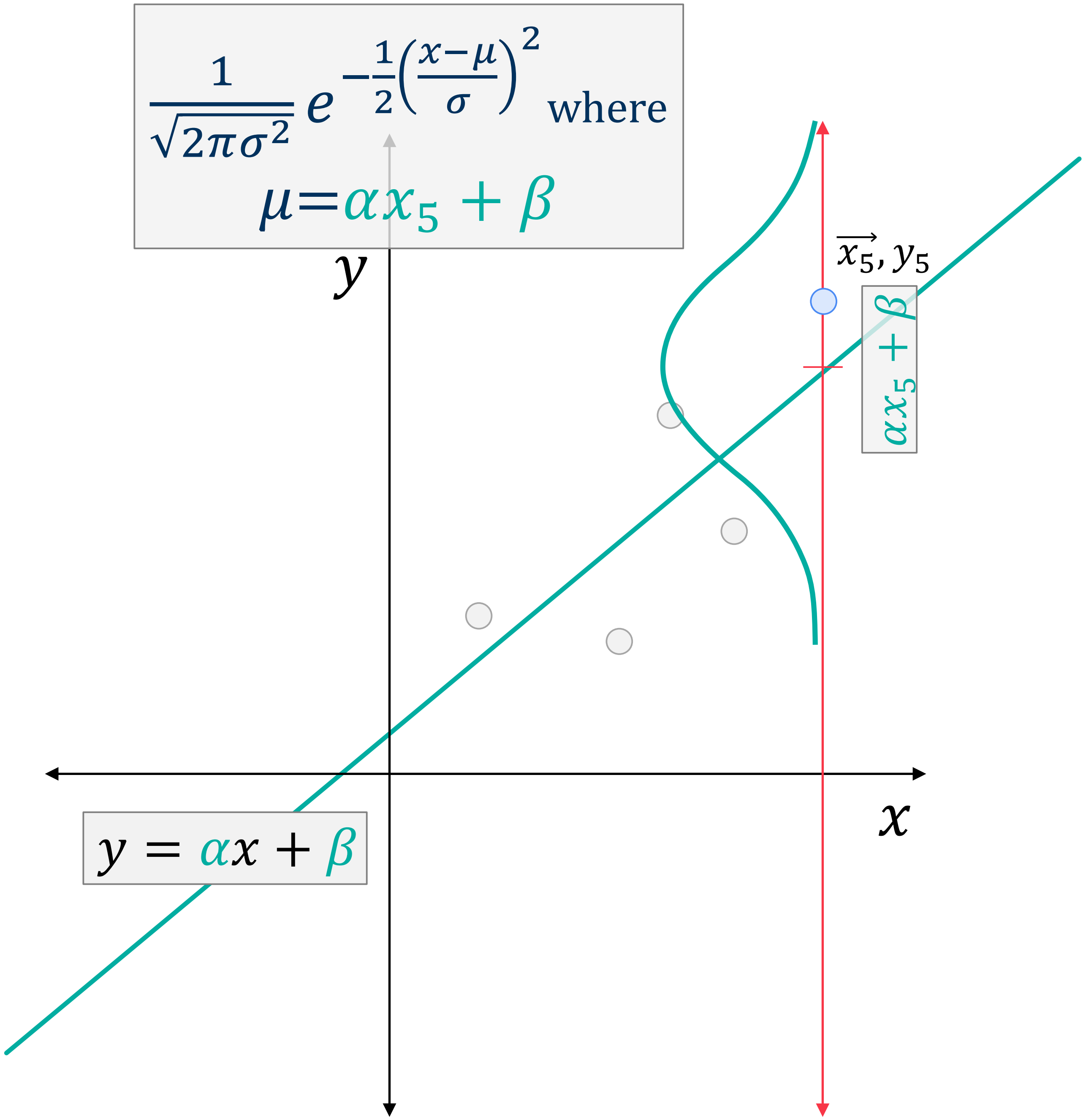
MLE example #2

Linear regression



MLE example #2

Linear regression

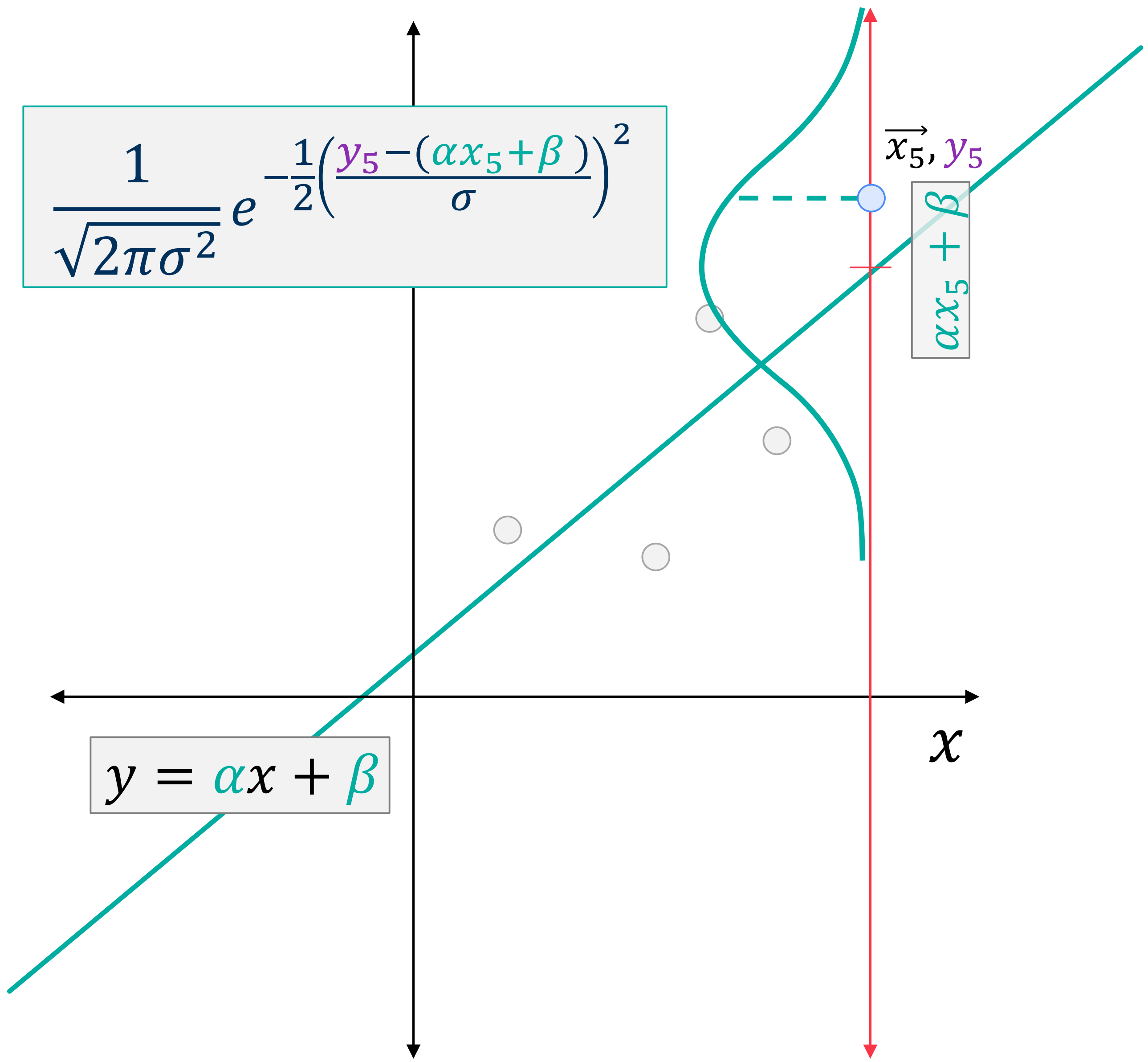


MLE example #2

Linear regression

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ where}$$

$$\mu = \alpha x_5 + \beta$$

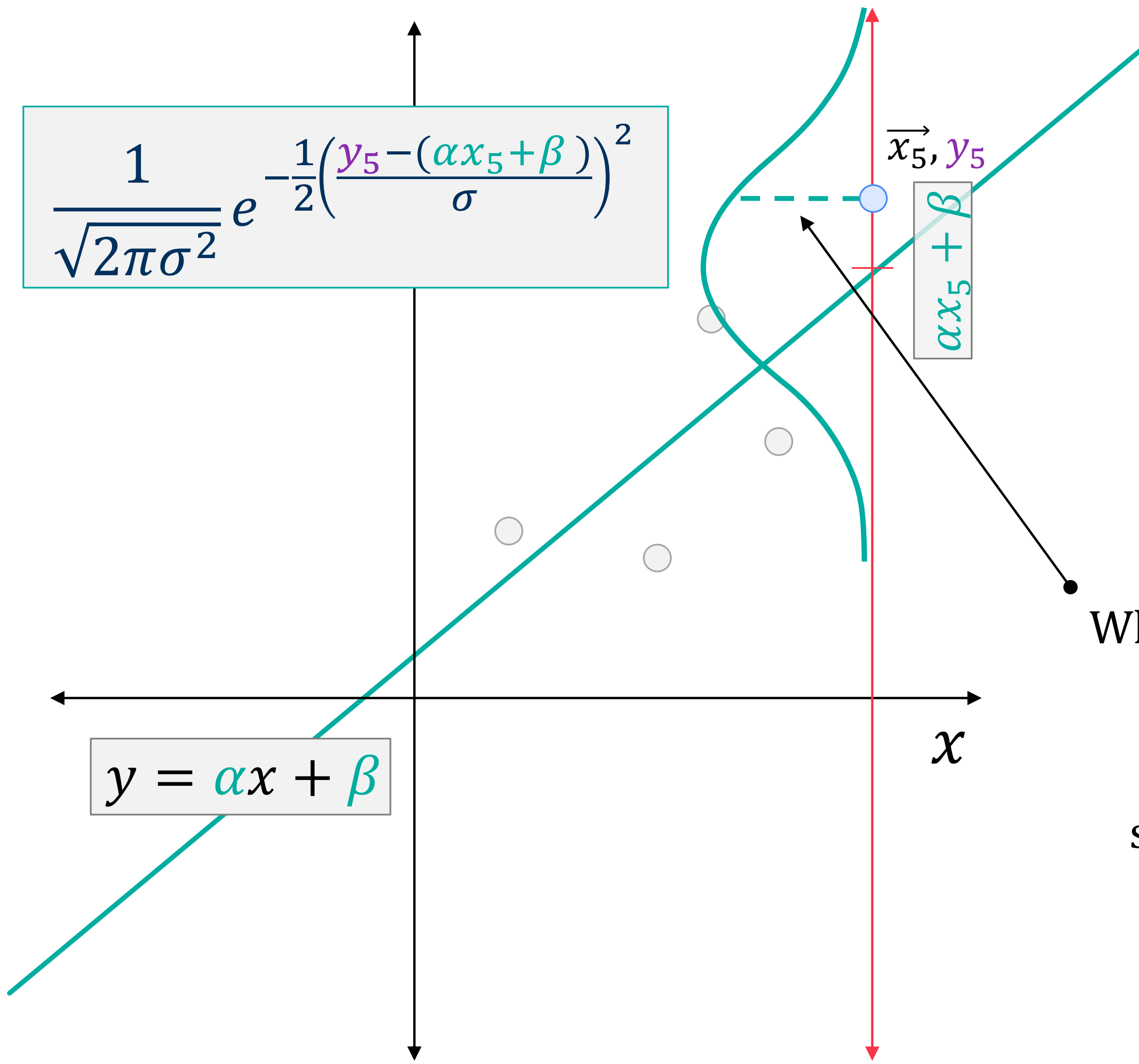


MLE example #2

Linear regression

$$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ where}$$

$$\mu = \alpha x_5 + \beta$$



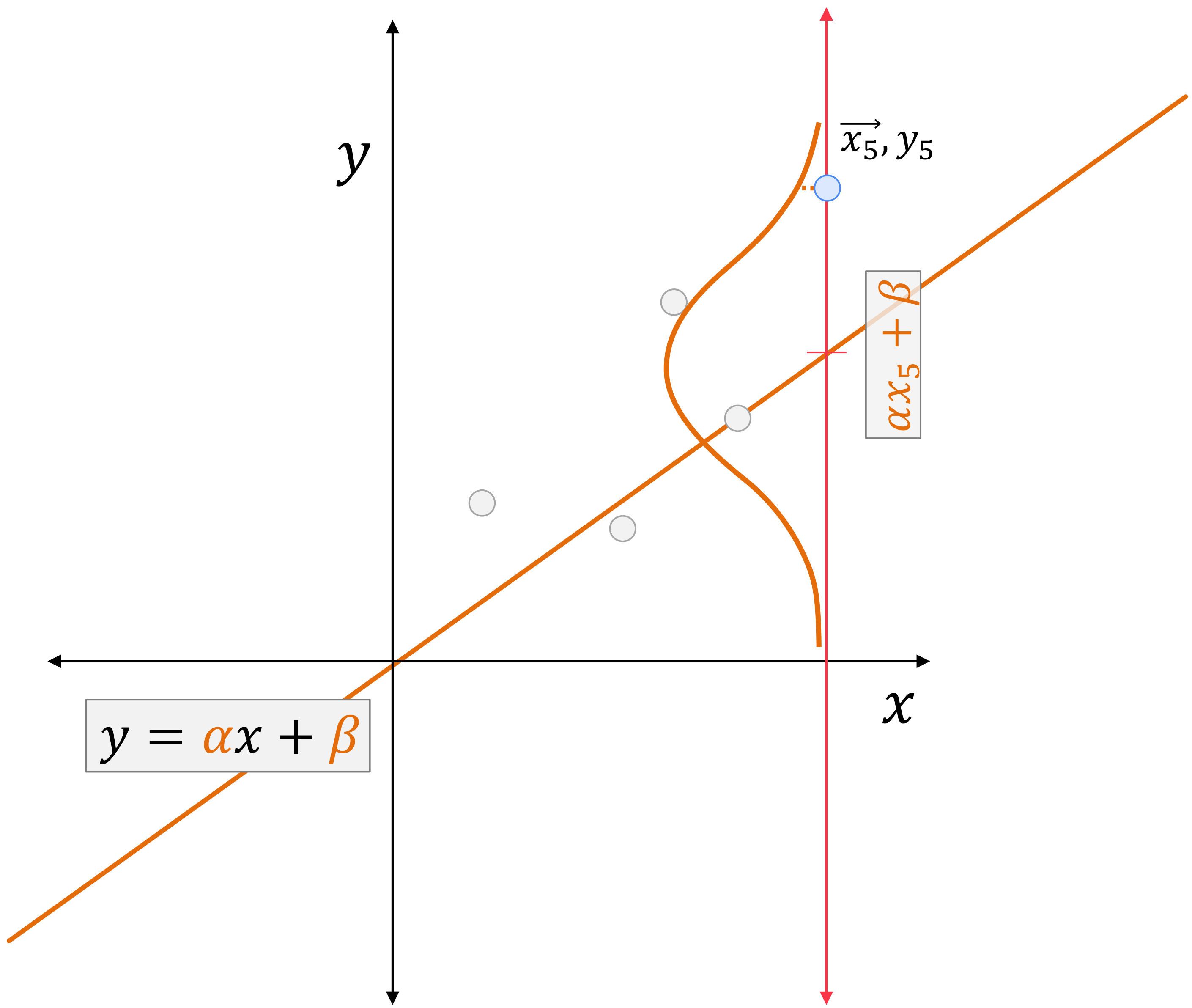
Which one do you more prefer

shorter

longer

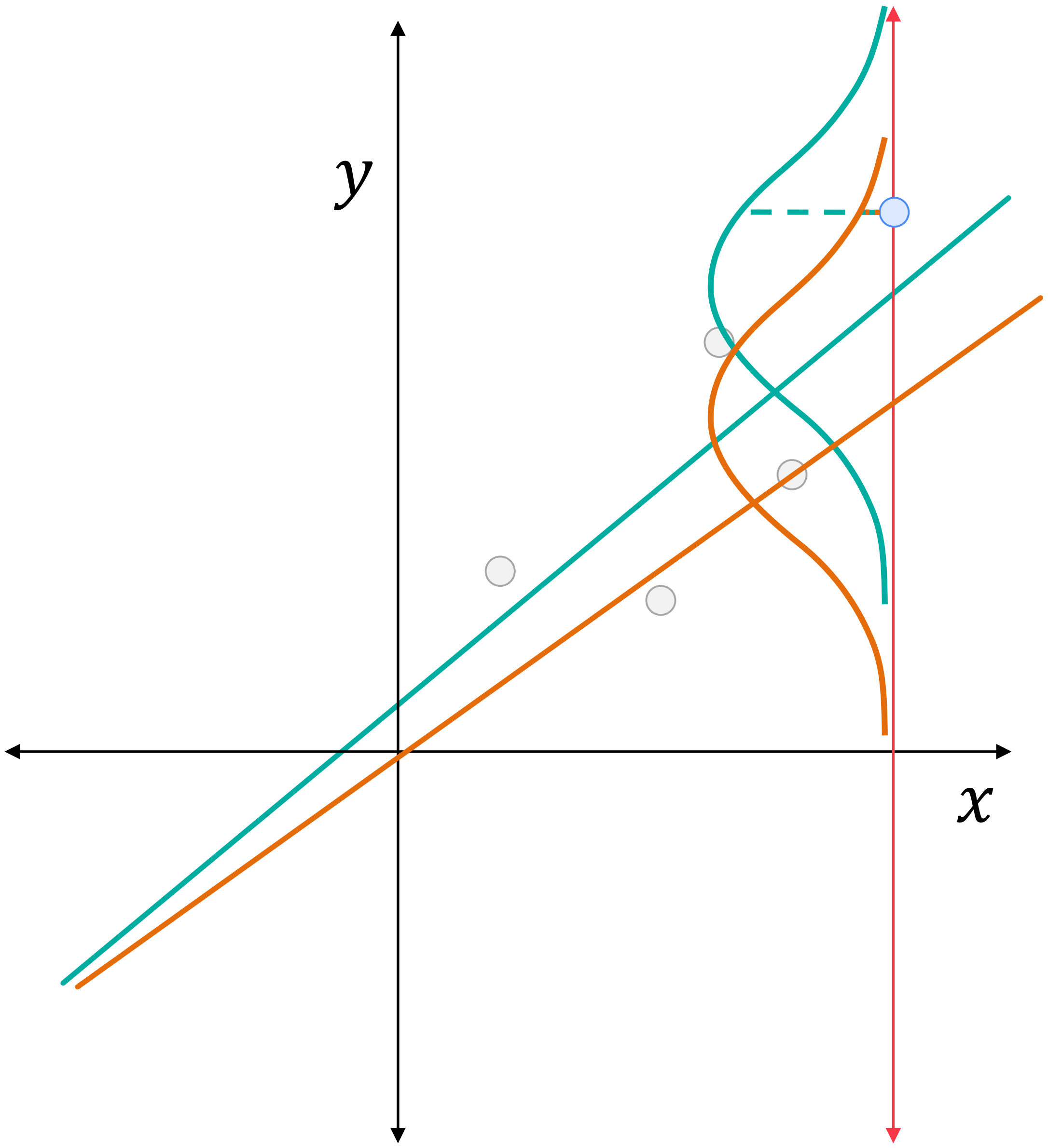
MLE example #2

Linear regression



MLE example #2

Linear regression



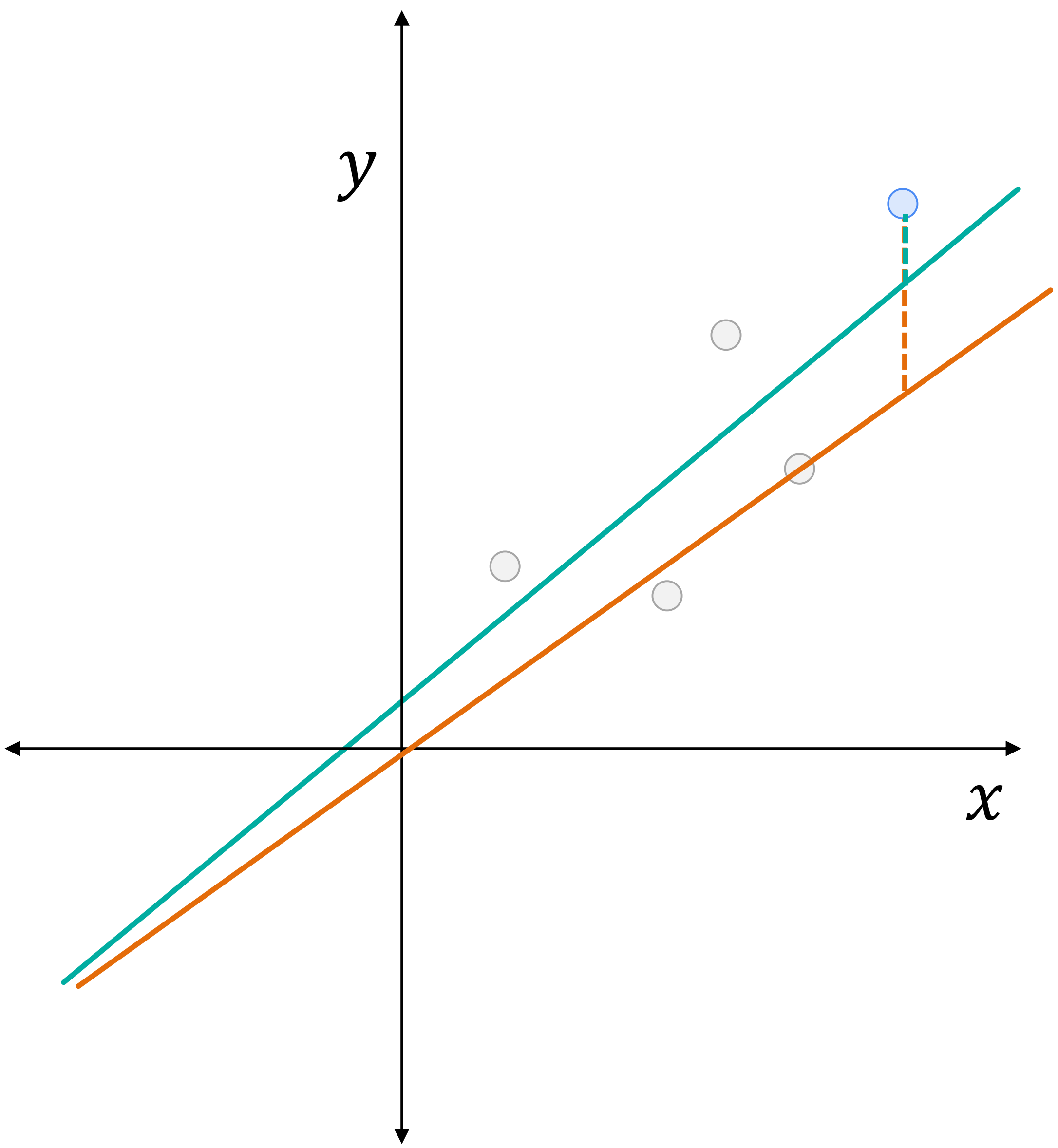
Which one do you more prefer

shorter

longer

MLE example #2

Linear regression

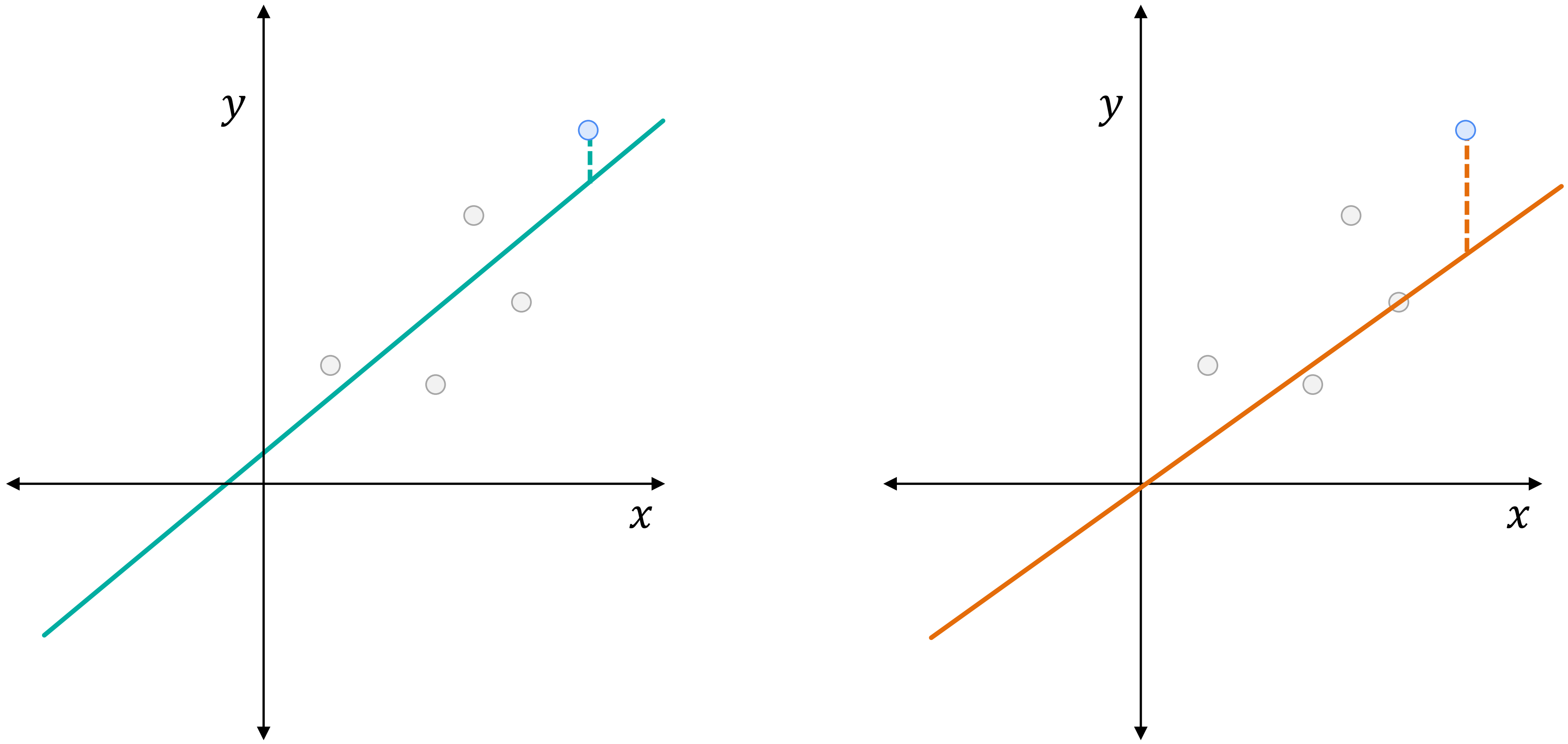


Which one do you more prefer ?

MLE example #2

Linear regression

- Which one do you more prefer?



MLE example #2

Linear regression

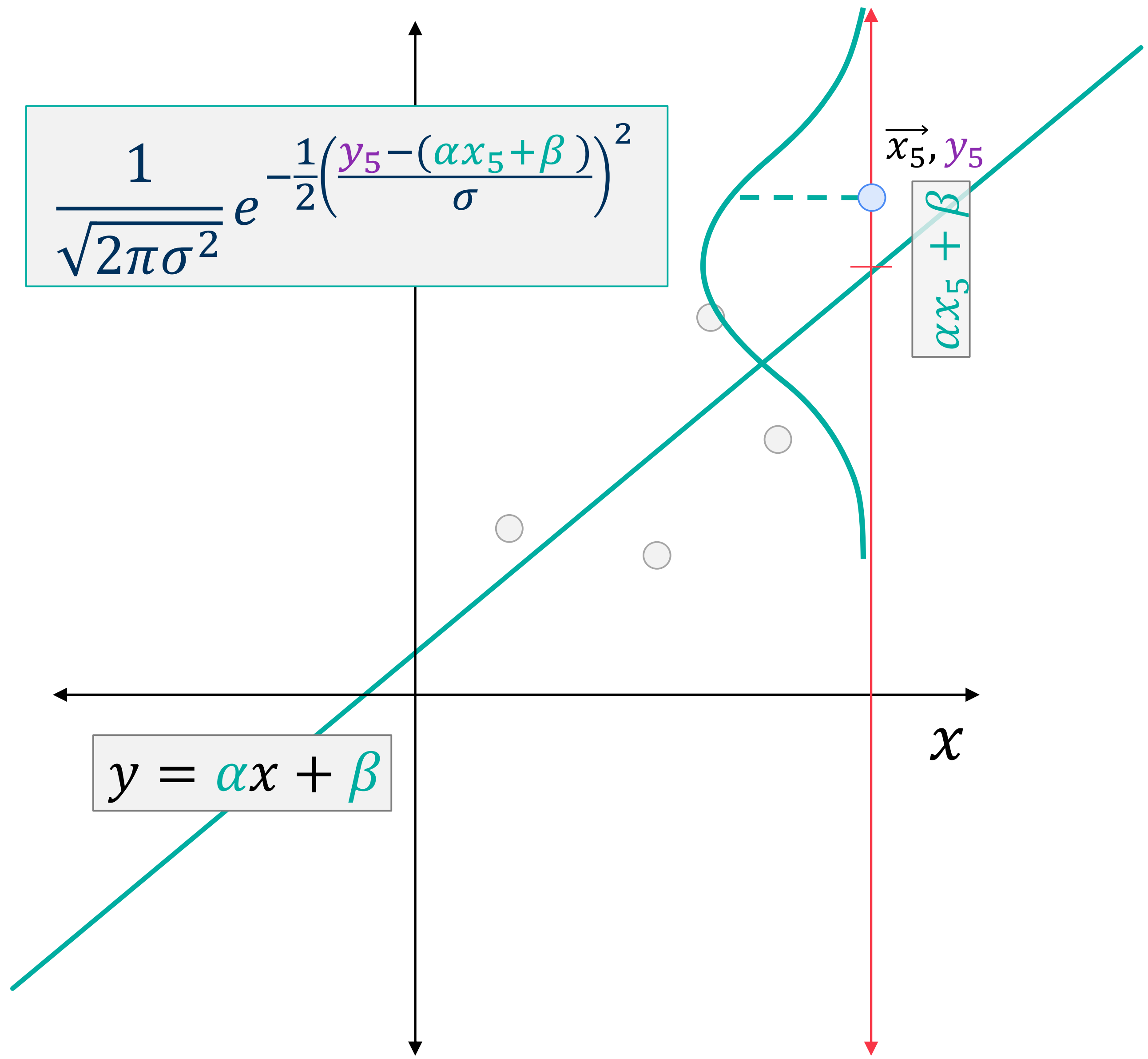
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ where}$$

$$\mu = \alpha x_5 + \beta$$

MLE

We need to find α and β maximizing

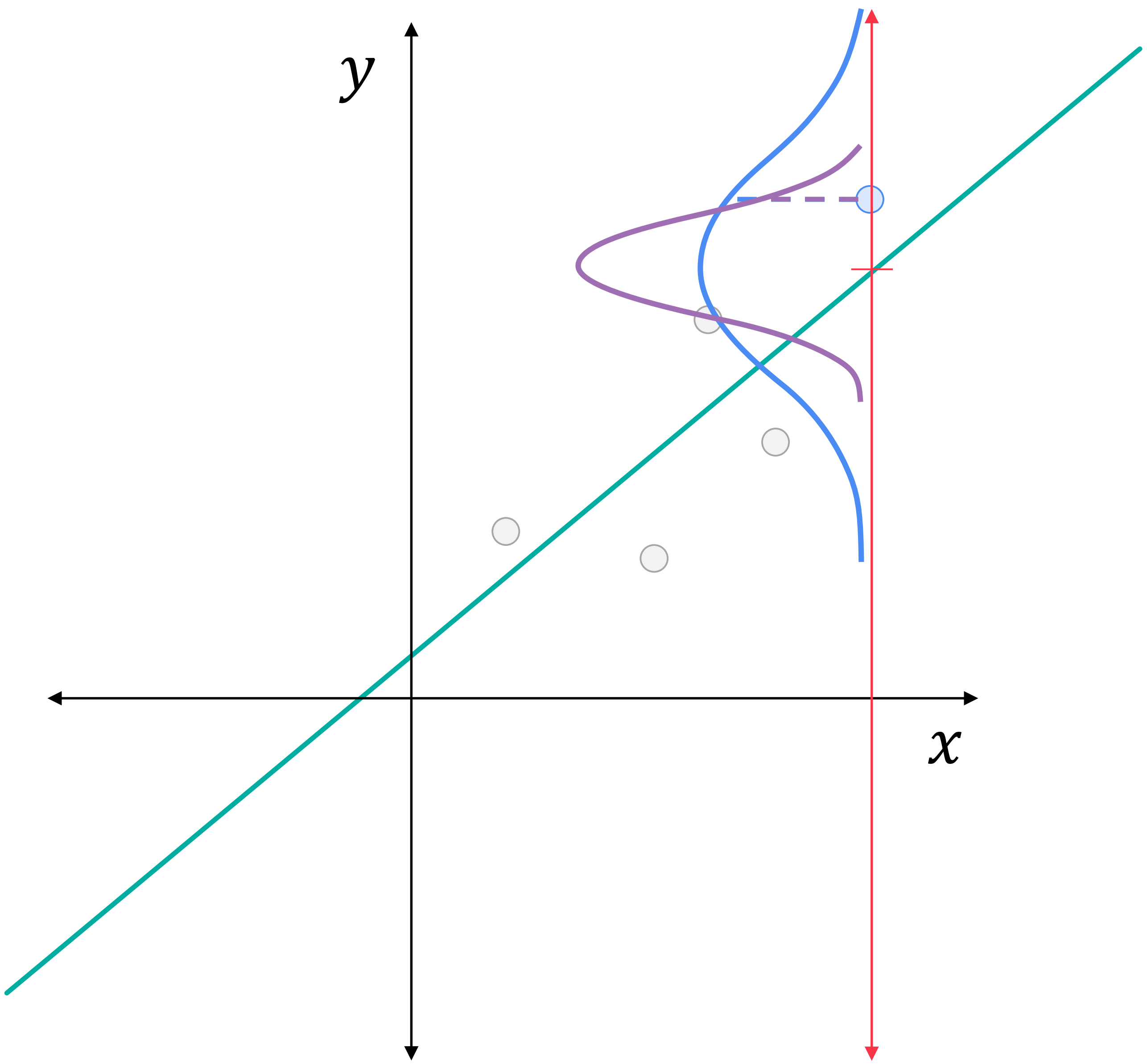
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_5-(\alpha x_5+\beta)}{\sigma}\right)^2}$$



MLE example #2

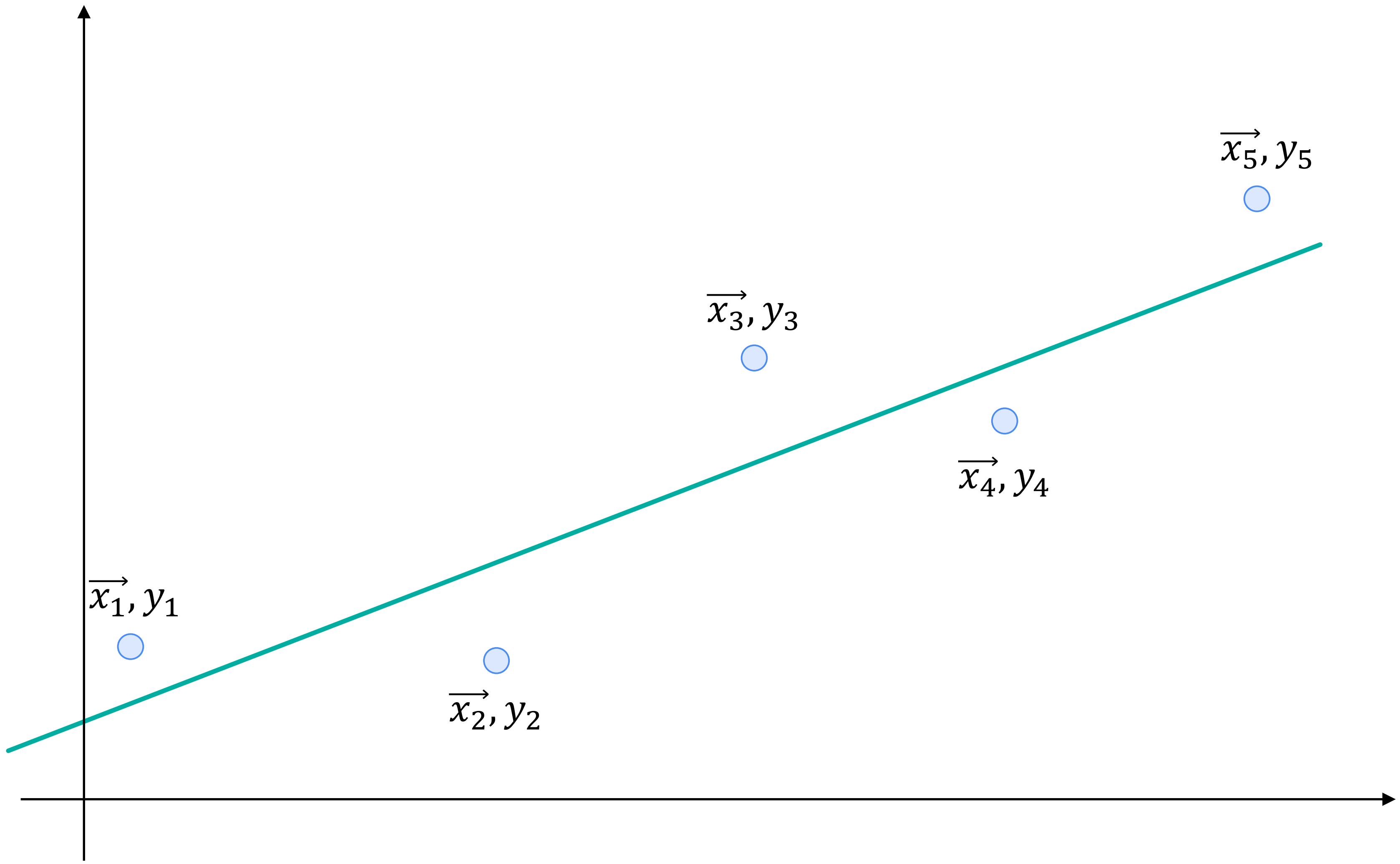
Linear regression

- Oh, σ doesn't matter!



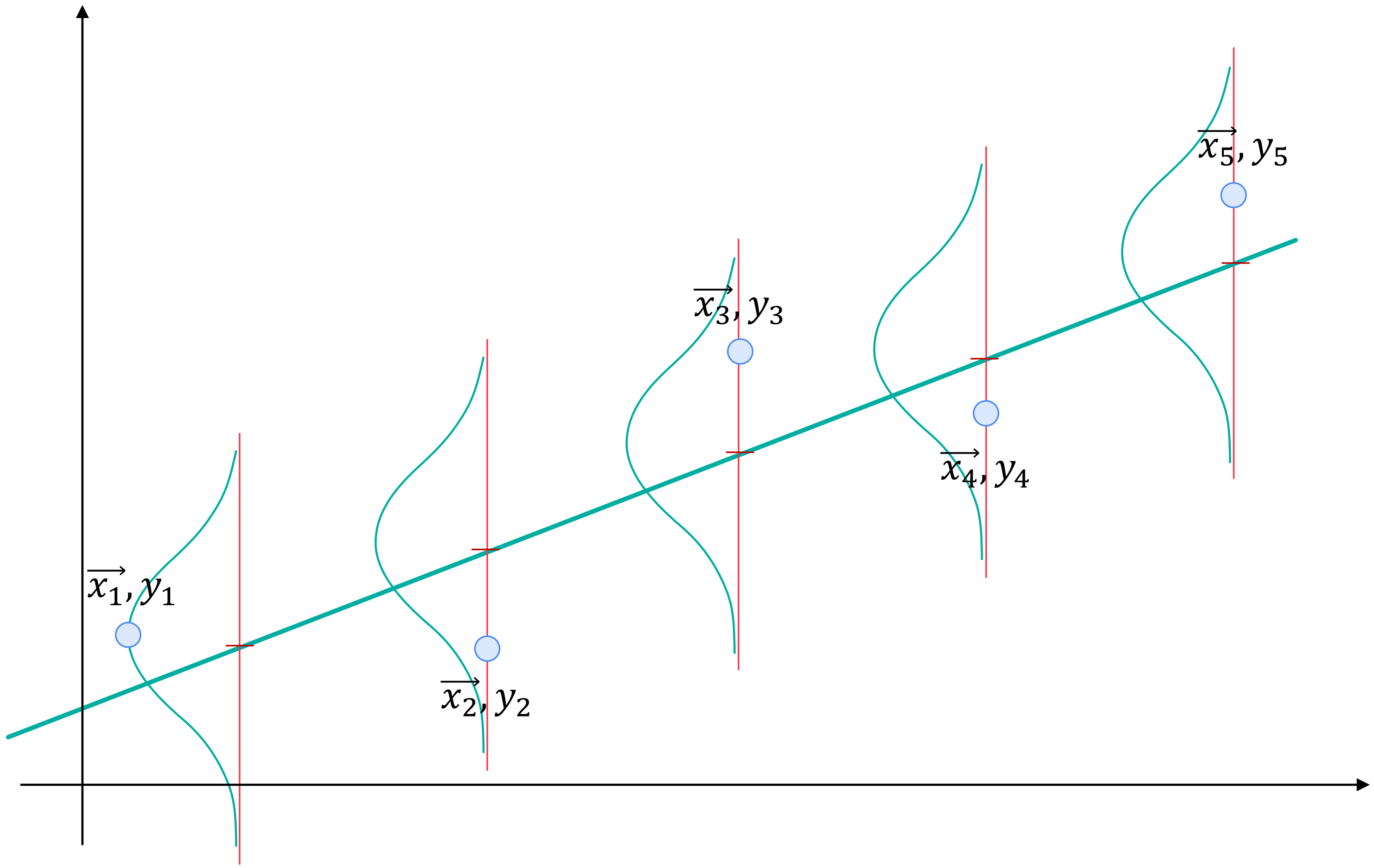
MLE example #2

Linear regression



MLE example #2

Linear regression



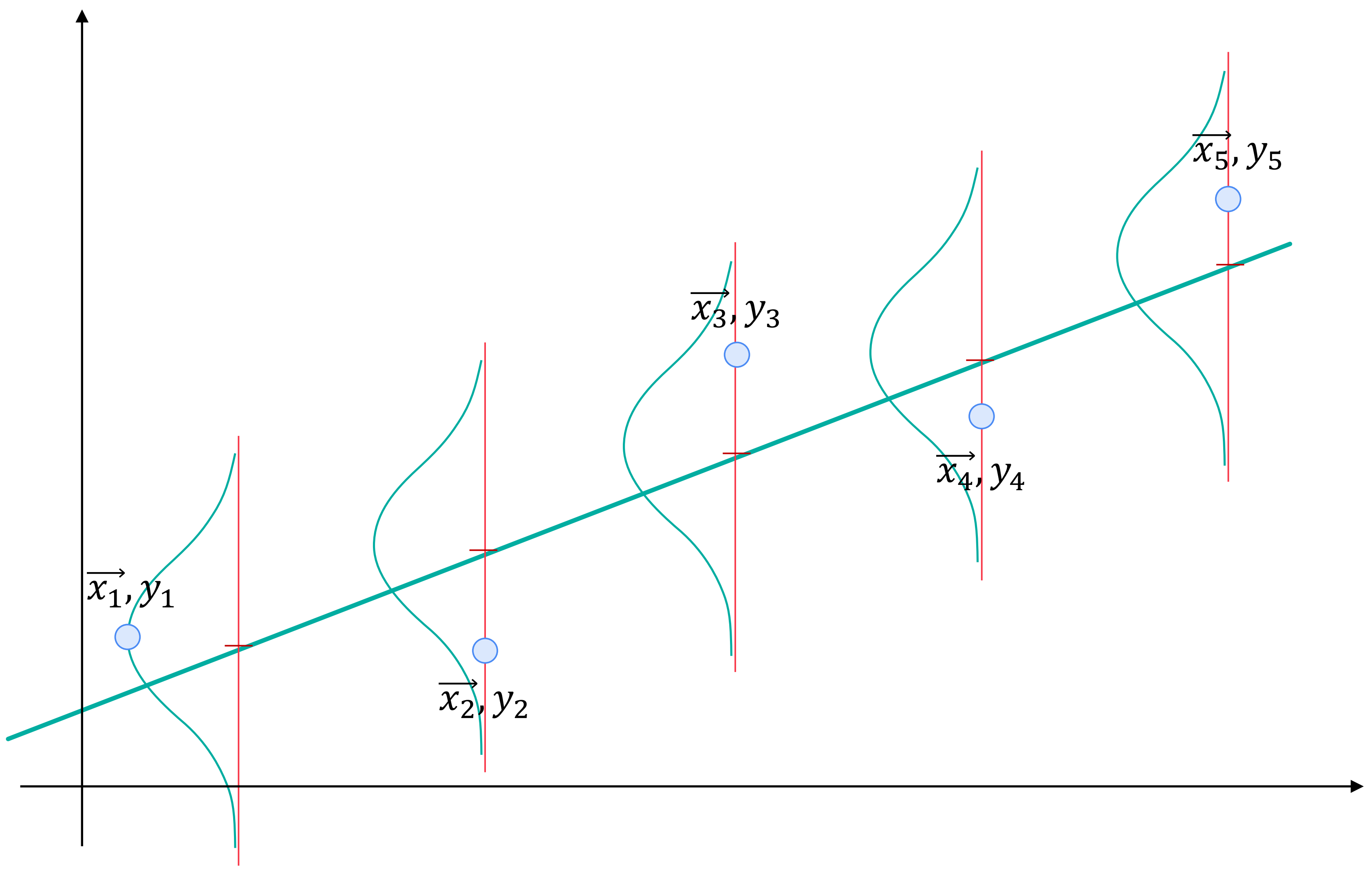
MLE example #2

Linear regression

We need to find α and β maximizing

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_5-(\alpha x_5+\beta)}{\sigma}\right)^2} \longrightarrow \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-(\alpha x_i+\beta)}{\sigma}\right)^2}$$

$1 \leq i \leq 5$



MLE example #2

Linear regression

- Likelihood function we need to maximize is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-(\alpha x_i+\beta\ )}{\sigma}\right)^2}$$

MLE example #2

Linear regression

- Likelihood function we need to maximize is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - (\alpha x_i + \beta)}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{y_i - (\alpha x_i + \beta)}{\sigma}\right)^2}$$

Log likelihood

$$L = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{y_i - (\alpha x_i + \beta)}{\sigma}\right)^2}\right)$$

MLE example #2

Linear regression

- Likelihood function we need to maximize is

$$L = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{y_i-(\alpha x_i+\beta)}{\sigma}\right)^2}\right)$$

$$= \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right) + \ln\left(\prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{y_i-(\alpha x_i+\beta)}{\sigma}\right)^2}\right)$$

$$= \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right) + \ln\left(e^{-\frac{1}{2}\left(\frac{y_1-(\alpha x_1+\beta)}{\sigma}\right)^2}\right) + \cdots + \ln\left(e^{-\frac{1}{2}\left(\frac{y_n-(\alpha x_n+\beta)}{\sigma}\right)^2}\right)$$

$$= \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right) + \ln\left(e^{-\frac{1}{2\sigma^2}(y_1-(\alpha x_1+\beta))^2}\right) + \cdots + \ln\left(e^{-\frac{1}{2\sigma^2}(y_n-(\alpha x_n+\beta))^2}\right)$$

$$= \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right) + -\frac{1}{2\sigma^2}(y_1-(\alpha x_1+\beta))^2\ln(e) + \cdots + -\frac{1}{2\sigma^2}(y_n-(\alpha x_n+\beta))^2\ln(e)$$

$$= \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\alpha x_i+\beta))^2$$

MLE example #2

Linear regression

- Likelihood function we need to maximize is

$$L = n\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\alpha x_i + \beta\ ))^2$$

$$\frac{\partial L}{\partial \alpha} = 0 \qquad\qquad \frac{\partial L}{\partial \beta} = 0$$

$$\boxed{SE_{Line} = (y_1 - (\alpha x_1 + \beta))^2 + (y_2 - (\alpha x_2 + \beta))^2 + \ldots + (y_n - (\alpha x_n + \beta))^2}$$

The thing we had tried to minimize last year ☺

This is why people say that

MLE is the basic & core technique of

the field of pattern recognition including

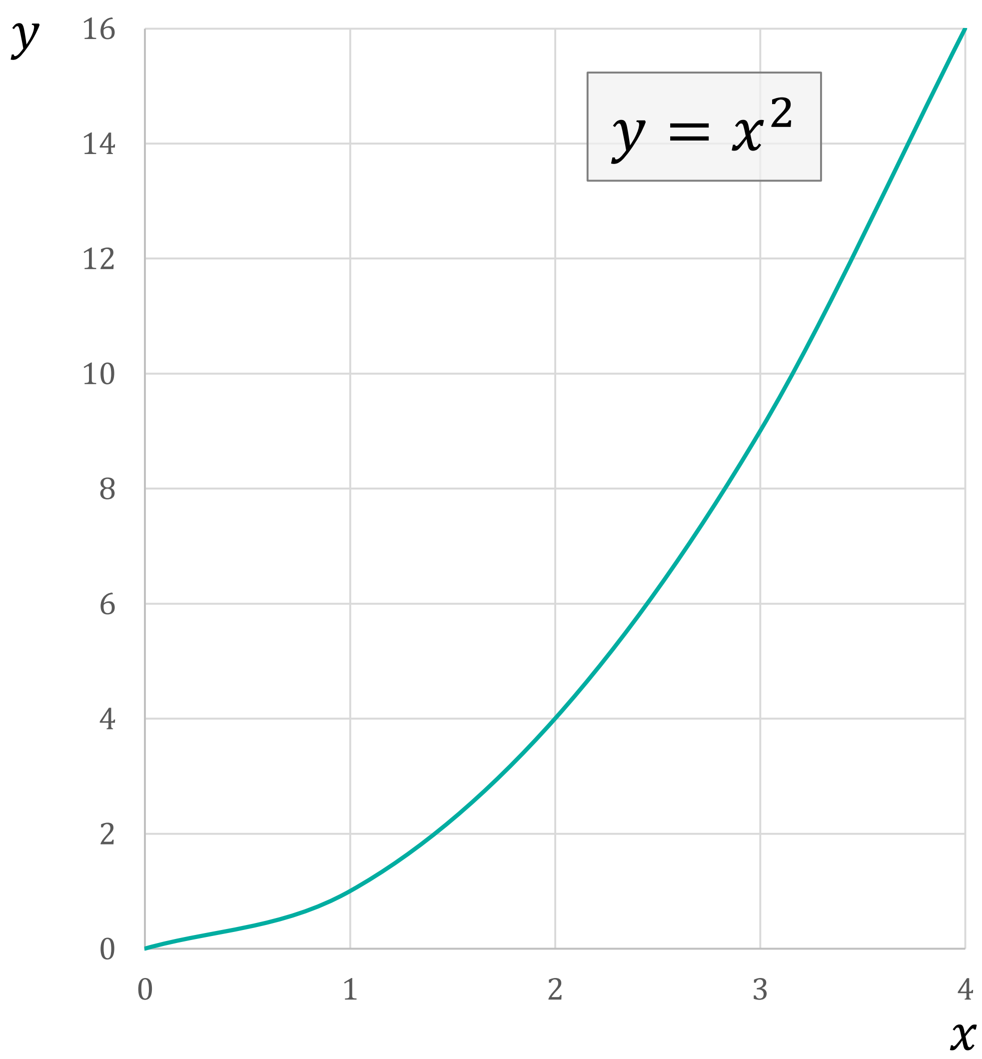
Deep learning, Support Vector Machine, Decision Trees,

Markov Random Field, Neural Networks, Linear Regression,

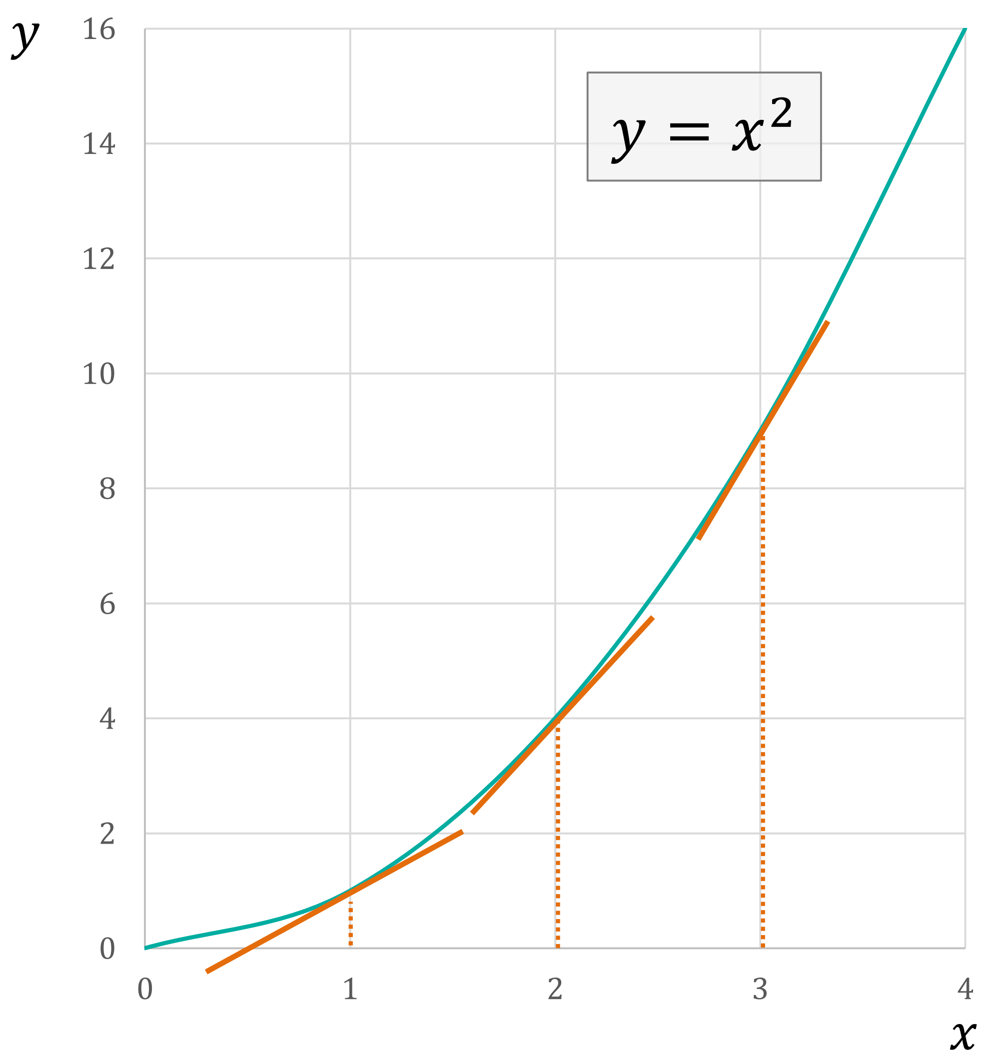
Logistic Regression, Maximum Entropy Model and etc.

Pieces I want to share with you

Derivative
도함수, 導函數



Derivative
도함수, 導函數



Derivative of $y = x^2$

- $\frac{dy}{dx} = 2x$ by using the 'power rule'
- the slope of the tangent line at any point along the curve
- the rate of change in y w.r.t x

Δx vs. dx vs. ∂x

Δx vs. dx vs. ∂x

Δx: delta x

dx: differential x

∂x: partial x

Δx vs. dx vs. ∂x

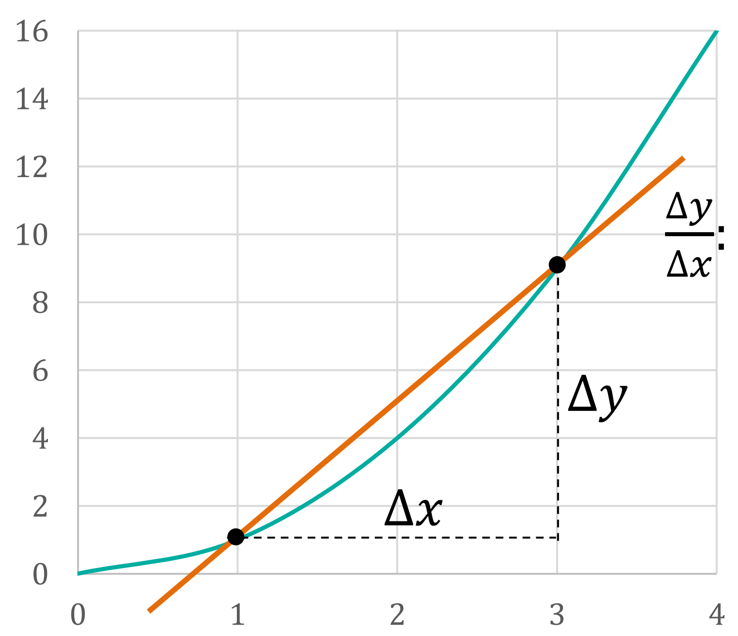
Δx: delta x

dx: differential x

∂x: partial x

About	Secant line (할선: 割線)

a line between two points representing
the rate of change between them



$\frac{\Delta y}{\Delta x}$: slope of the secant line

Δx vs. dx vs. ∂x

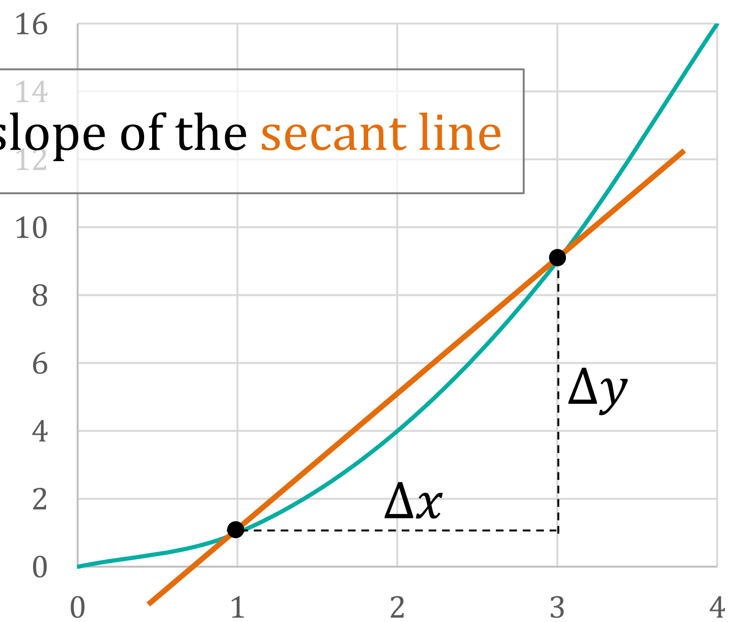
Δx: delta x	dx: differential x	∂x: partial x

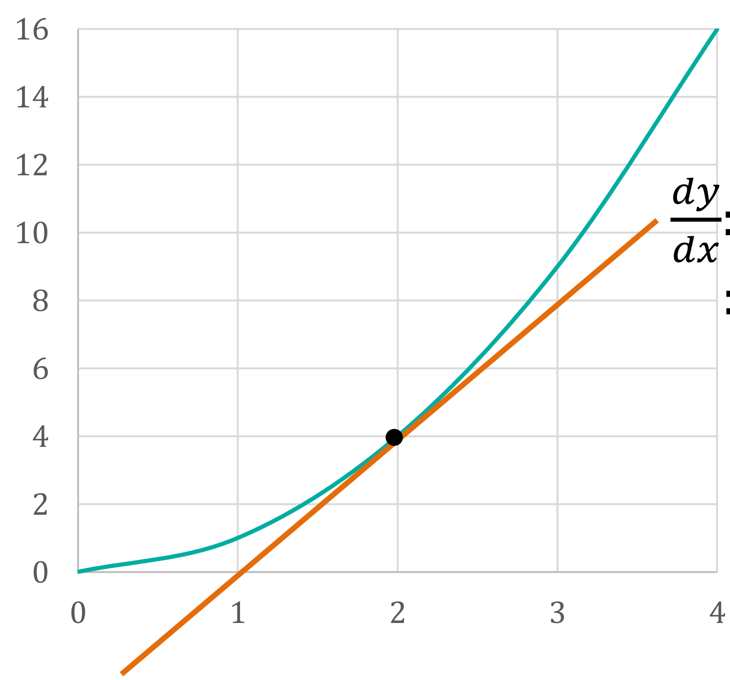
About	Secant line (할선: 割線)

About	Tangent line (접선: 接線)

a line to one point, representing
the rate of infinitely small (infinitesimal) change

$\frac{\Delta y}{\Delta x}$: slope of the secant line





$\frac{dy}{dx}$: slope of the tangent line
: derivative (도함수, 導函數) of y w.r.t x

Δx vs. dx vs. ∂x

Δx: delta x

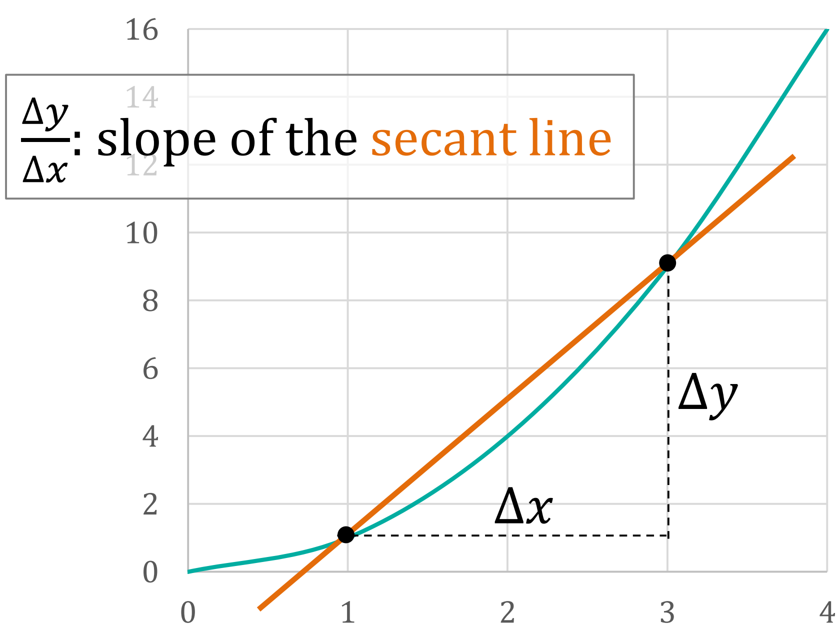
dx: differential x

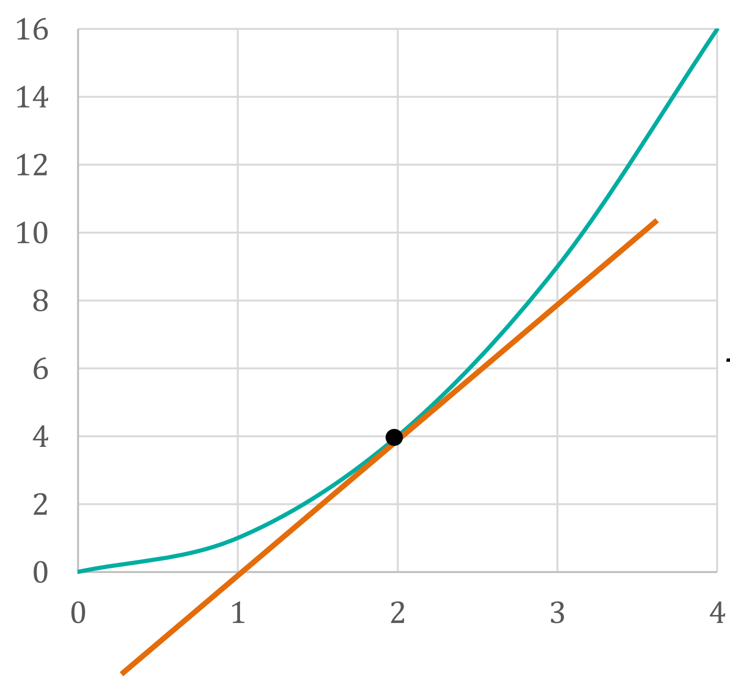
∂x: partial x

About	Secant line (할선: 割線)

About	Tangent line (접선: 接線)

a line to one point, representing
the rate of infinitely small (infinitesimal) change





$f(x)$

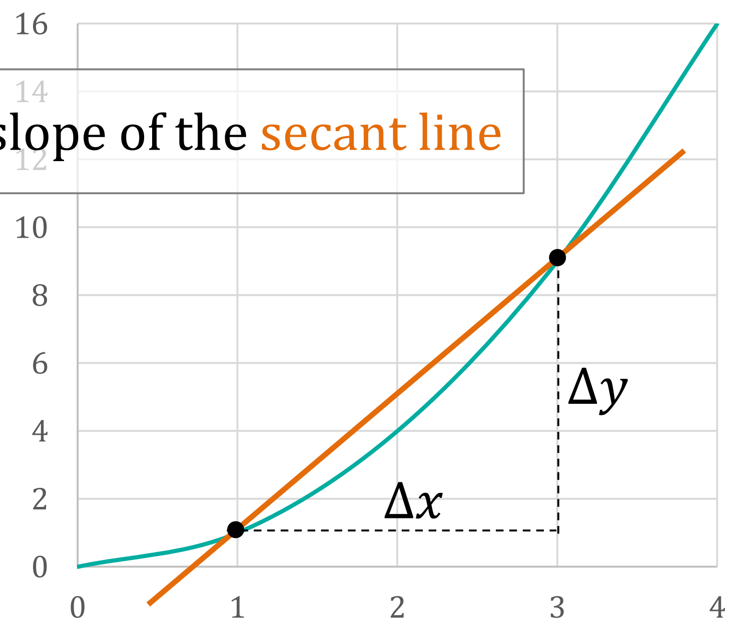
$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Δx vs. dx vs. ∂x

Δx: delta x	dx: differential x	∂x: partial x
About Secant line (할선: 割線)	**About** Tangent line (접선: 接線)	**About** Tangent line to a partial derivative

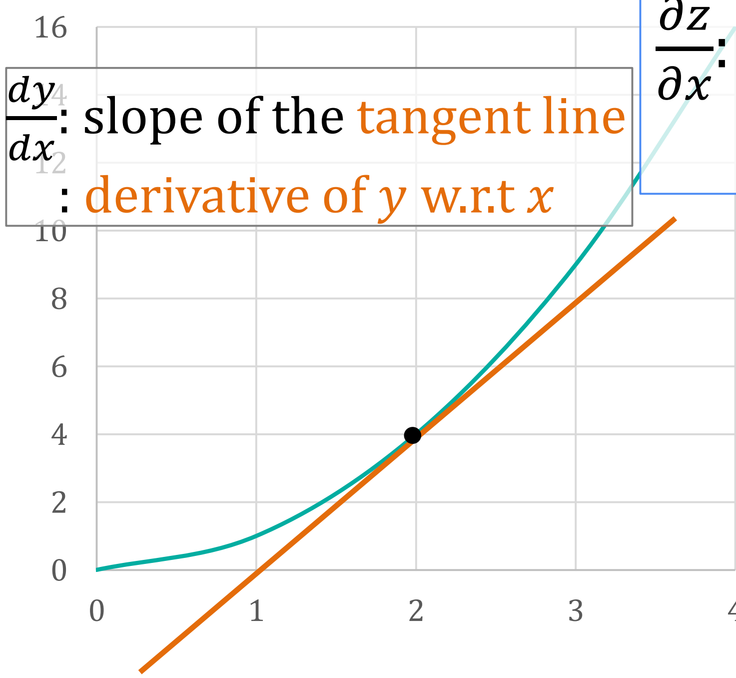
A rate of change or derivative in one direction, holding a number of other directions constant

$\frac{\Delta y}{\Delta x}$: slope of the secant line

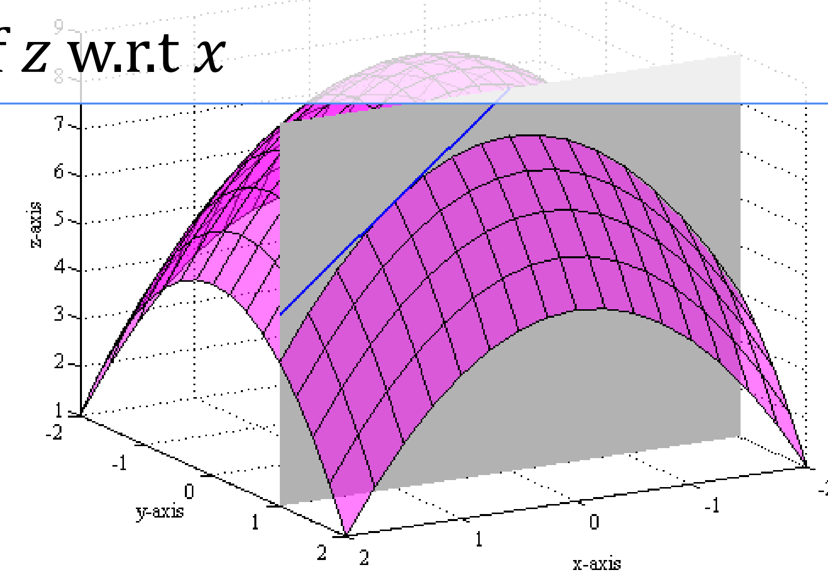


$\frac{dy}{dx}$: slope of the tangent line

: derivative of y w.r.t x



$\frac{\partial z}{\partial x}$: partial derivative (편도함수 偏導函數) of z w.r.t x



$$f'(x) \text{ vs. } \frac{dy}{dx} \text{ vs. } \frac{\partial y}{\partial x}$$

$f'(x)$ vs. $\frac{dy}{dx}$ vs. $\frac{\partial y}{\partial x}$

Inventor of infinitesimal calculus

Isaac Newton

Gottfried Leibniz

$$f(x) = x^2 + x$$

$$f'(x) = 2x + 1 \longleftrightarrow \frac{d(x^2 + x)}{dx} = \frac{d}{dx}(x^2 + x) = 2x + 1$$

$f'(x)$ vs. $\frac{dy}{dx}$ vs. $\frac{\partial y}{\partial x}$

$$\frac{d}{dx}f(x)$$

Total derivative

$$\frac{\partial}{\partial x}f(x, y, z)$$

Partial derivative

Derivative of $\ln(x)$

Natural constant e
(데이터사이언스개론)

Compound interest

Bernoulli

$$\left(1+\frac{1}{n}\right)^{n}$$

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} = 2.718281\ldots = e$$

Interest rate: 100%

1 year

1 → 2 ⇢ $(1+1)^1$

Interest rate: 50% Interest rate: 50%

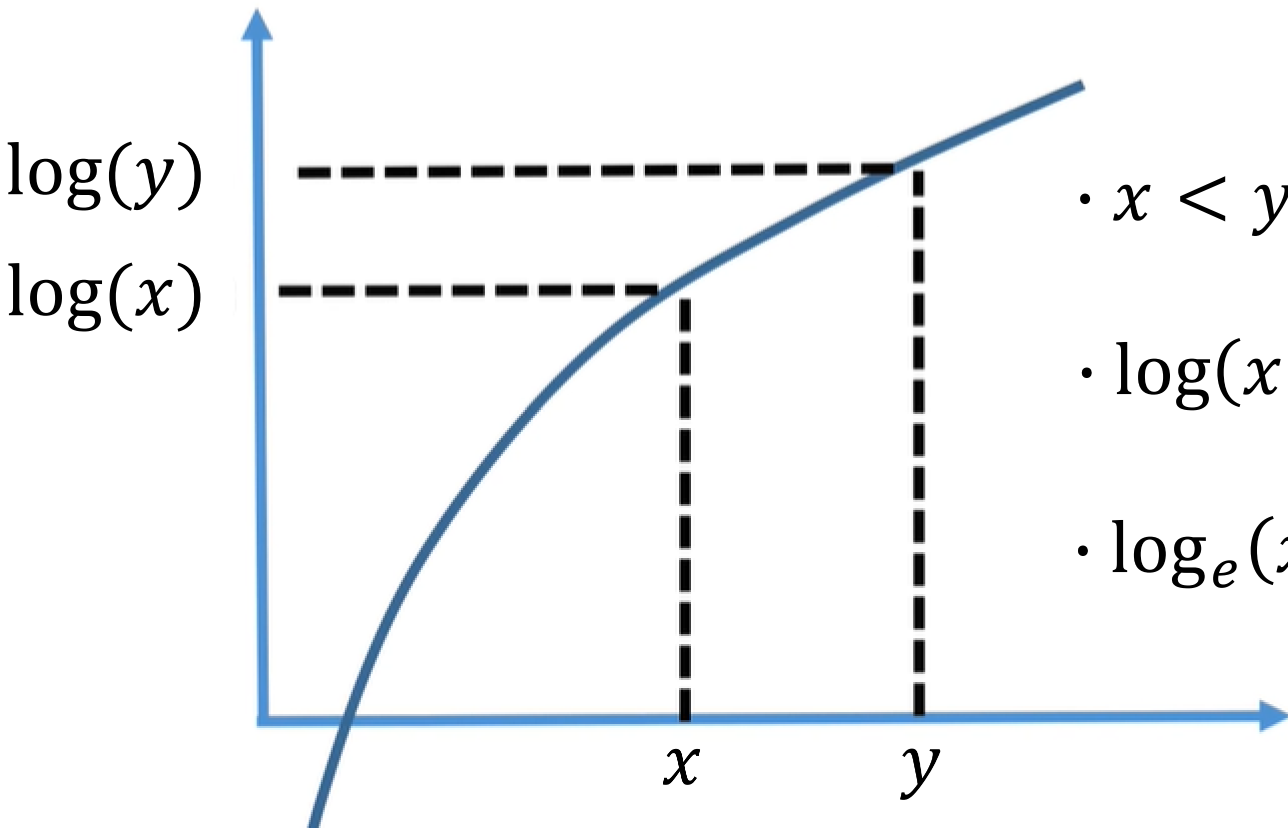
1 → → 2.25 ⇢ $\left(1+\frac{1}{2}\right)^2$

→ → → → ⇢ $\left(1+\frac{1}{4}\right)^4$

…

every day! ⇢ $\left(1+\frac{1}{365}\right)^{365}$

How to maximize the probability of our observation
(데이터사이언스응용)



- $x < y \implies \log(x) < \log(y)$
- $\log(x \times y \times z) = \log(x) + \log(y) + \log(z)$
- $\log_e(x) = \ln(x) \longrightarrow \boxed{(\ln(x))' = \frac{1}{x}}$

Derivative of $\ln(x)$

- Let $f(x) = \ln(x)$.
- $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{ln(x+h) - ln(x)}{h}$$

$$= \lim_{h \to 0} \frac{ln\left(\frac{x+h}{x}\right)}{h}$$

$$= \lim_{h \to 0} \frac{1}{h} ln\left(\frac{x+h}{x}\right)$$

$$= \lim_{h \to 0} \left(ln\left(\frac{x+h}{x}\right)^{\frac{1}{h}}\right)$$

$$= ln\left(\lim_{h \to 0} \left(\frac{x+h}{x}\right)^{\frac{1}{h}}\right)$$

$$= ln\left(\lim_{h \to 0} \left(1 + \frac{h}{x}\right)^{\frac{1}{h}}\right)$$

Derivative of $\ln(x)$

- Let $f(x) = \ln(x)$.
- $f'(x) = ln\left(\lim_{h\to 0}\left(1 + \frac{h}{x}\right)^{\frac{1}{h}}\right)$

Derivative of $\ln(x)$

- Let $f(x) = \ln(x)$.
- $f'(x) = ln\left(\lim_{h\to 0}\left(1 + \frac{h}{x}\right)^{\frac{1}{h}}\right)$

Definition of e

$$\boxed{\lim_{t\to\infty}\left(1 + \frac{1}{t}\right)^{t} = e}$$

$$\downarrow$$

- Let $u = \frac{1}{t}$ then $\begin{cases} t \to \infty \\ u \to 0 \end{cases}$

$$\downarrow$$

$$\boxed{\lim_{u\to 0}(1 + u)^{\frac{1}{u}} = e}$$

Derivative of $\ln(x)$

- Let $f(x) = \ln(x)$.
- $f'(x) = \boxed{ln\left(\lim_{h\to 0}\left(1+\frac{h}{x}\right)^{\frac{1}{h}}\right)}$

$\downarrow$

- Let $u = \frac{h}{x}$ then $\begin{cases} h \to 0 \\ u \to 0 \end{cases}$

$\downarrow$

$$\boxed{ln\left(\lim_{u\to 0}(1+u)^{\frac{1}{ux}}\right)}$$

$$\boxed{= ln\left(\lim_{u\to 0}\left((1+u)^{\frac{1}{u}}\right)^{\frac{1}{x}}\right)}$$

$$\boxed{= ln\left(\lim_{u\to 0}(1+u)^{\frac{1}{u}}\right)^{\frac{1}{x}}}$$

Definition of e

$$\boxed{\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^{t} = e}$$

$\downarrow$

- Let $u = \frac{1}{t}$ then $\begin{cases} t \to \infty \\ u \to 0 \end{cases}$

$\downarrow$

$$\boxed{\lim_{u\to 0}(1+u)^{\frac{1}{u}} = e}$$

Derivative of $\ln(x)$

- Let $f(x) = \ln(x)$.
- $f'(x) = \boxed{ln\left(\lim_{h\to 0}\left(1+\frac{h}{x}\right)^{\frac{1}{h}}\right)}$

- Let $u = \frac{h}{x}$ then $\begin{cases} h \to 0 \\ u \to 0 \end{cases}$

$$\boxed{ln\left(\lim_{u\to 0}(1+u)^{\frac{1}{ux}}\right)}$$

$$\boxed{= ln\left(\lim_{u\to 0}\left((1+u)^{\frac{1}{u}}\right)^{\frac{1}{x}}\right)}$$

$$\boxed{= ln\left(\lim_{u\to 0}(1+u)^{\frac{1}{u}}\right)^{\frac{1}{x}}}$$

Definition of e

$$\boxed{\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^{t} = e}$$

- Let $u = \frac{1}{t}$ then $\begin{cases} t \to \infty \\ u \to 0 \end{cases}$

$$\boxed{\lim_{u\to 0}(1+u)^{\frac{1}{u}} = e}$$

$$\boxed{= ln(e)^{\frac{1}{x}}}$$

$$\boxed{= \frac{1}{x}}$$

과제 1. Proof of the derivative of sigmoid function

$$f(x) = \frac{1}{1+e^{-x}} \longrightarrow f'(x) = f(x)(1 - f(x))$$

- 제출 일자: 2020.10.06 오후 11:59
- 제출 파일 형식: PDF
- 제출 방법
  - e-강의동 > 데이터사이언스응용 > 과제 1. Proof of the derivative of sigmoid function

Chain rule

IF a variable z depends on the variable y, which itself depends on the variable x,

(i.e., y and z are therefore dependent variables)

THEN z, via the intermediate variable of y, depends on x as well.

The chain rule then states,

$$\boxed{\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}}$$

Chain rule with a simple example

$$z = 6\ln(1 - x) \qquad \frac{dz}{dx} = ?$$

Chain rule with a simple example

$$z = 6\ln(1 - x)$$

$$\frac{dz}{dx} = ?$$

$$y = 1 - x$$

$$z = 6\ln(y)$$

y depends on x

z, via the intermediate variable y, depends on x

Chain rule with a simple example

$$z = 6\ln(1 - x)$$

$$\frac{dz}{dx} = ?$$

$$y = 1 - x$$

$$z = 6\ln(y)$$

y depends on x

z, via the intermediate variable y, depends on x

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}$$

Chain rule with a simple example

$z = 6\ln(1 - x)$

$$\frac{dz}{dx} = ?$$

$\boxed{\frac{dy}{dx} = -1}$ $y = 1 - x$

$z = 6\ln(y)$ $\boxed{\frac{dz}{dy} = \frac{6}{y}}$

y depends on x

z, via the intermediate variable y, depends on x

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}$$

Chain rule with a simple example

$$z = 6\ln(1 - x)$$

$$\frac{dz}{dx} = ?$$

$$\boxed{\frac{dy}{dx} = -1} \quad y = 1 - x$$

$$z = 6\ln(y) \quad \boxed{\frac{dz}{dy} = \frac{6}{y}}$$

y depends on x

z, via the intermediate variable y, depends on x

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}$$

$$\frac{dz}{dx} = \frac{6}{y} \times -1 = \frac{6}{1-x} \times -1 = -\frac{6}{1-x}$$

Proof of chain rule

- Given $z = f(g(x))$

$$\frac{dz}{dx} = \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f(g(x))}{h}\right)$$

$$= \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f(g(x))}{h} \times \frac{g(x+h)-g(x)}{g(x+h)-g(x)}\right)$$

$$= \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f(g(x))}{g(x+h)-g(x)} \times \frac{g(x+h)-g(x)}{h}\right)$$

$$= \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f(g(x))}{g(x+h)-g(x)}\right) \times \lim_{h\to 0}\left(\frac{g(x+h)-g(x)}{h}\right)$$

$$= \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f\big(g(x)\big)}{g(x+h)-g(x)}\right) \times g'(x)$$

Proof of chain rule

- Given $z = f(g(x))$

$$\frac{dz}{dx} = \lim_{h\to 0}\left(\frac{f\big(g(x+h)\big)-f\big(g(x)\big)}{g(x+h)-g(x)}\right) \times g'(x)$$

$$= \lim_{k\to 0}\left(\frac{f(g(x)+k)-f\big(g(x)\big)}{k}\right) \times g'(x)$$

$$= f'(g(x)) \times g'(x)$$

Let $k = g(x+h) - g(x)$

$$\begin{cases} k \to 0 \text{ where } h \to 0 \\ g(x+h) = g(x) + k \end{cases}$$

Thank you