AI VIET NAM
@aivietnam.edu.vn

Module 02 – Exercise Class

NAÏVE BAYES CLASSIFIERS

Nguyen Quoc Thai

Objectives

Review

- Probability
- Conditional Probability
- Total Probability Theorem
- Bayes' Rule

Bayes Classifiers

- Naïve Bayes Classifier
- Exercise
- Implementation

AI VIET NAM
@aivietnam.edu.vn

Outline

SECTION 1
Review

SECTION 2
Naïve Bayes Classifier

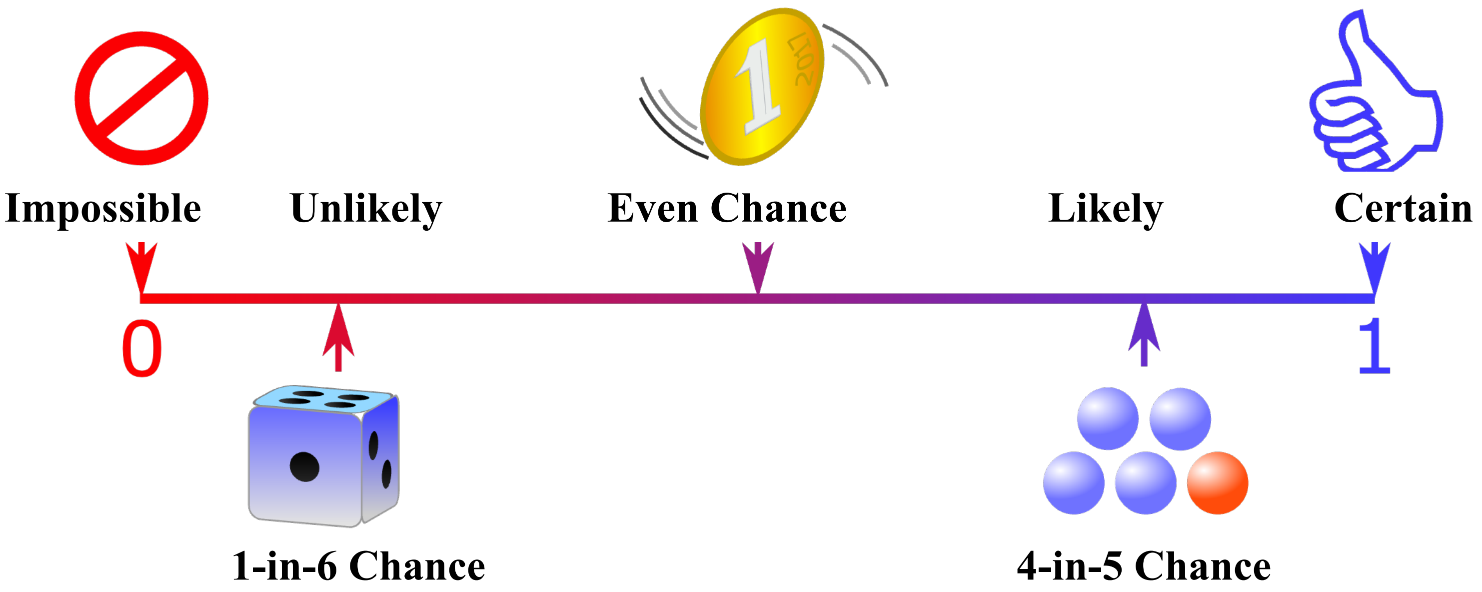
SECTION 3
Exercise

SECTION 4
Implementation

Review

Probability

- Measure to the likelihood of an event occurring



Review

Classical Probability

$$P(A) = \frac{number\ of\ favorable\ outcomes}{total\ number\ of\ possible\ outcomes} = \frac{n_A}{n_\Omega}$$

Example

What is the probability of rolling a number is even on a regular dice?

- There are 6 faces on a fair die, numbered 1 to 6 => n(Ω) = 6
- A : "even number" => A = {2, 4, 6} => n(A) = 3

=> P(A) = 3/6 = 0.5

Review

Geometric Probability

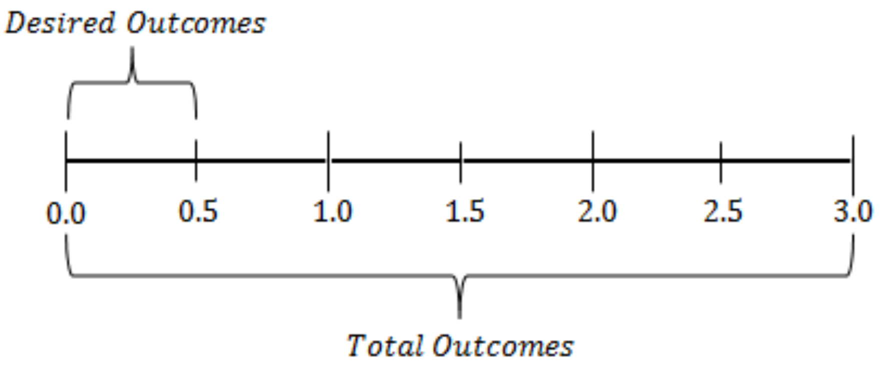
$$P(A) = \frac{measure\ of\ domain\ A}{measure\ of\ domain\ \Omega}$$

1-D Geometric probability

X is a random real number between 0 and 3. What is the probability X is closer to 0 than it is to 1?

=> A: “X is closer to 0 than to 1”

=> Measure: length in this 1D case: $P(A) = \frac{length\ of\ segment\ where\ 0<X<0.5}{length\ of\ segment\ where\ 0<X<3} = \frac{0.5}{3} = \frac{1}{6}$



Review

Geometric Probability

$$P(A) = \frac{measure\ of\ domain\ A}{measure\ of\ domain\ \Omega}$$

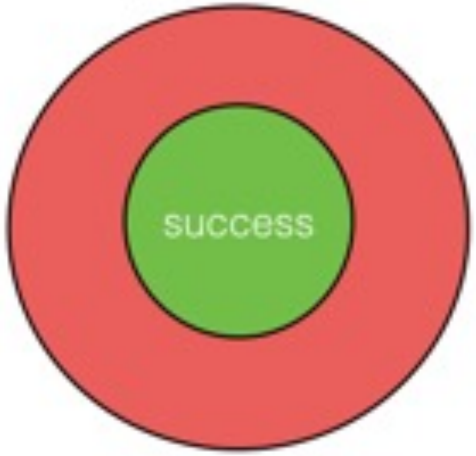
➢ 2-D Geometric probability

A dart is thrown at a circular dartboard such that it will land randomly over the area of the dartboard. What is the probability that it lands closer to the center "success" than to the edge?

=> A: "closer to center than edge"

=> Measure: area in this 2D case:

$$P(A) = \frac{area\ of\ desired\ outcomes}{area\ of\ total\ outcomes} = \frac{\frac{\pi r^2}{4}}{\pi r^2} = \frac{1}{4}$$



Review

Rules of Probability

- Rule 1: For any event A, $0 \le P(A) \le 1$; $P(A^c) = 1 - P(A)$
- Rule 2: S - Sample space => $P(S) = 1$
- Rule 3: Addition rule: $P(A+B) = P(A) + P(B) - P(AB)$

  If A, B are mutually exclusive => $P(AB) = 0$
- Rule 4: Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

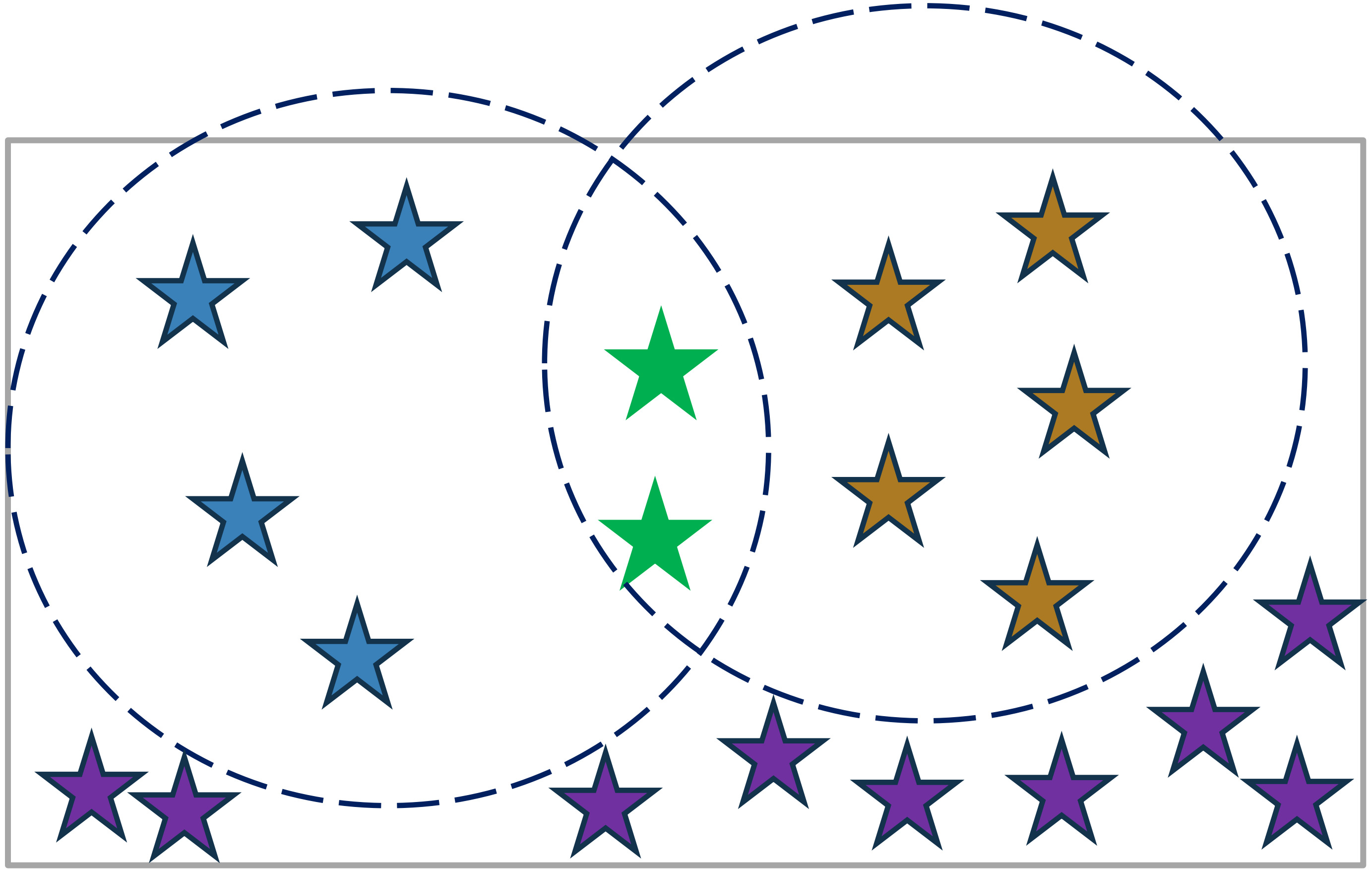
- Rule 5: Multiplication rule:

$$P(AB) = P(A).P(B|A) = P(B).P(A|B)$$

- Rule 6: Independent events: $P(AB) = P(A).P(B)$

Review

Practice

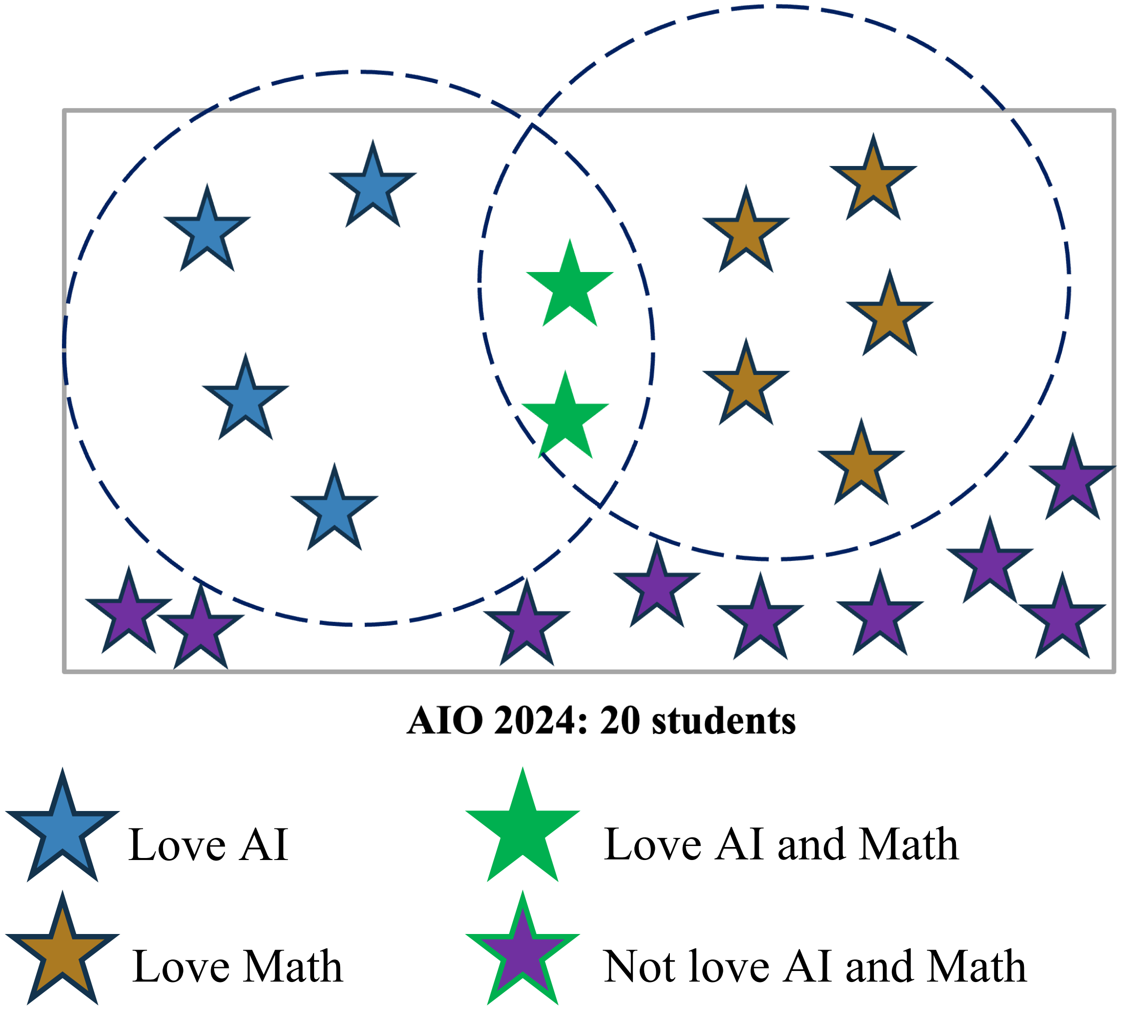


AIO 2024: 20 students

	Love AI	4
	Love Math	5
	Love AI and Math	2
	Not love AI and Math	9

Review

Practice



AIO 2024: 20 students

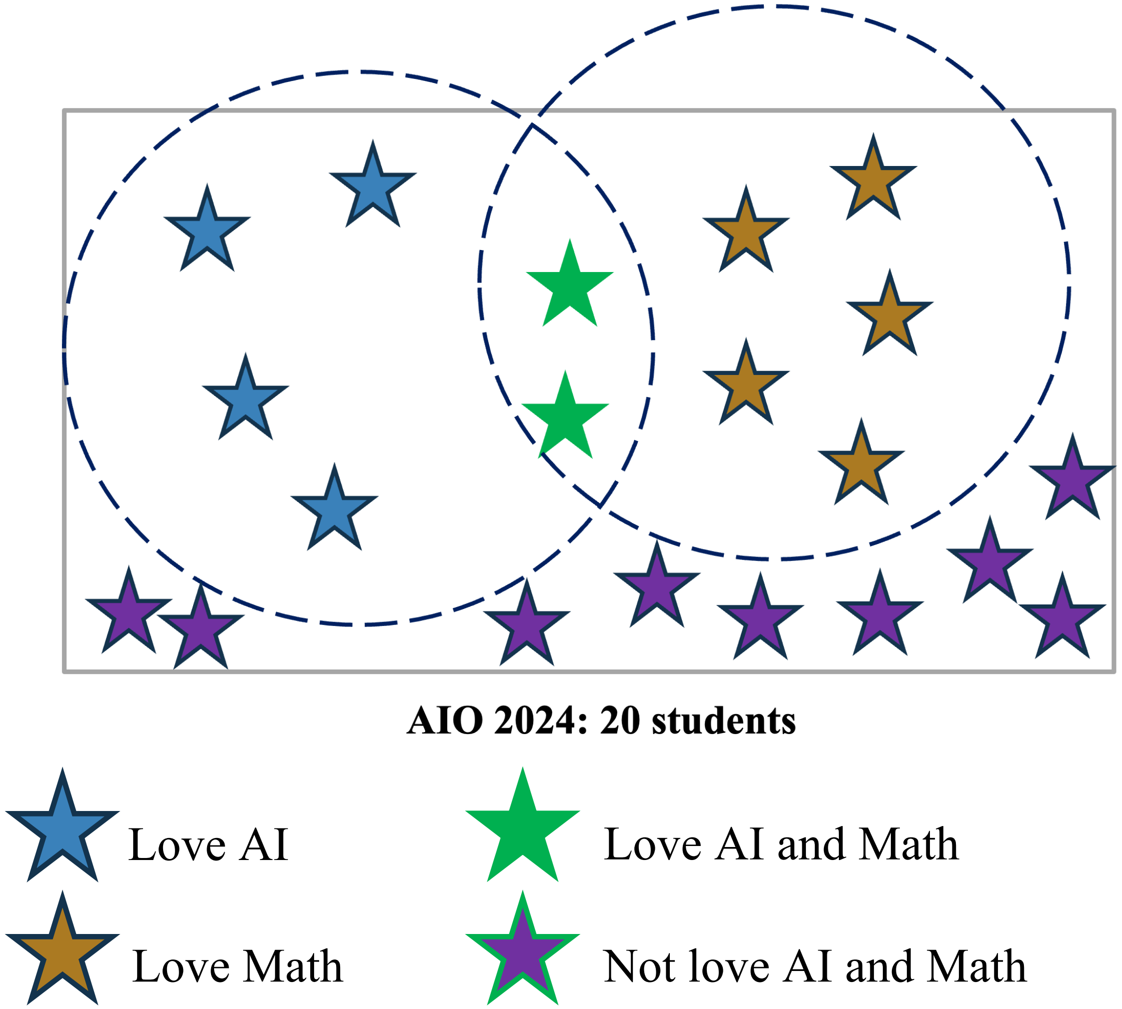
	Love AI	Not love AI
Love Math	2	5
Not love Math	4	9

The probability of meeting someone who love AI and Math in AIO 2023

$$p(\text{love AI and Math}) = \frac{2}{20} = 0.1 = 10\%$$

Review

Practice



AIO 2024: 20 students

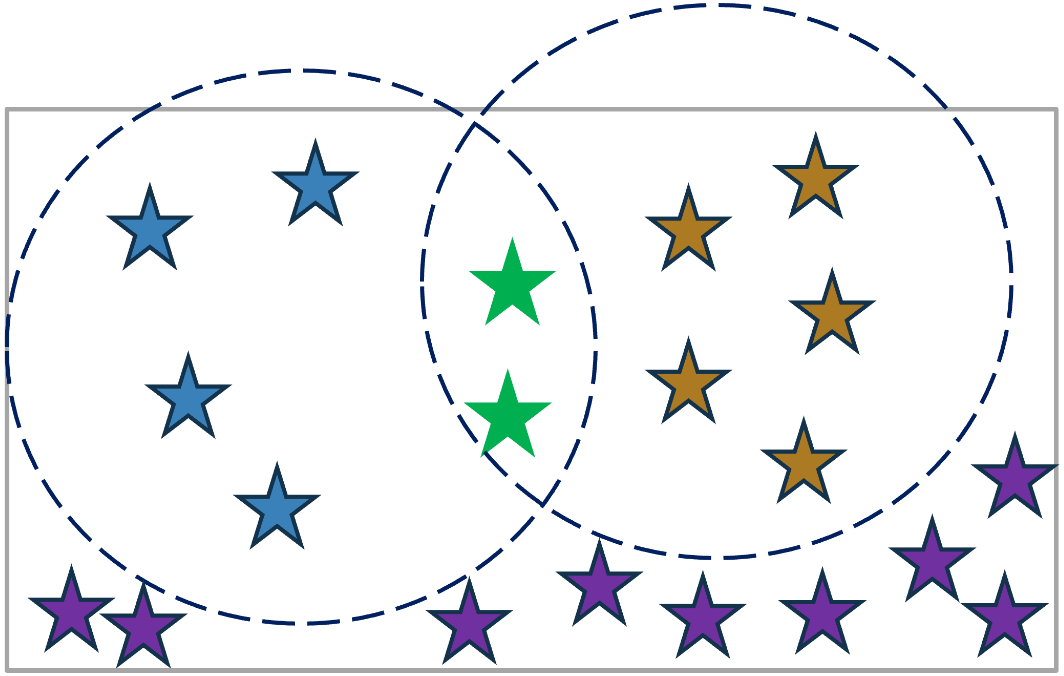
	Love AI	Not love AI	Total
Love Math	2 p = 2/20	5 p = 5/20	7 p = 7/20
Not love Math	4 p = 4/20	9 p = 9/20	13 p = 13/20

The probability that someone loves Math regardless how they feel about AI

The probability that someone does not loves Math regardless how they feel about AI

Review

Practice



AIO 2024: 20 students

Love AI
Love AI and Math
Love Math
Not love AI and Math

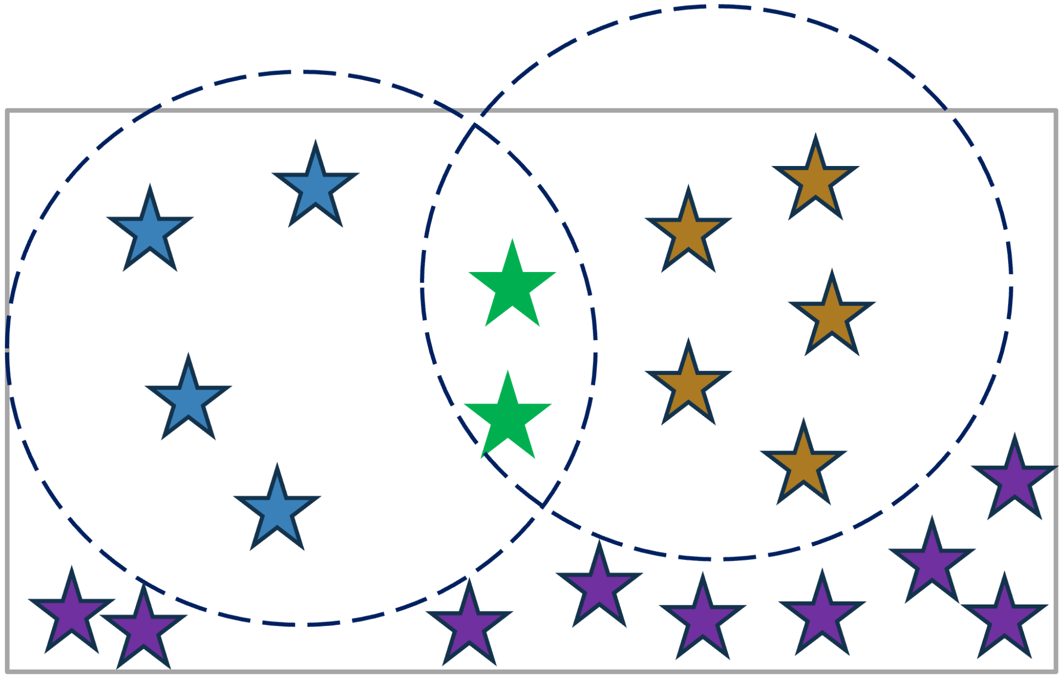
	Love AI	Not love AI	Total
Love Math	2 p = 2/20	5 p = 5/20	7 p = 7/20
Not love Math	4 p = 4/20	9 p = 9/20	13 p = 13/20
Total	6 P=6/20	14 P=14/20	

The probability that someone loves AI regardless how they feel about Math

The probability that someone does not loves AI regardless how they feel about Math

Review

Practice



AIO 2024: 20 students

	Love AI	Not love AI	Total
Love Math	2 p = 2/20	5 p = 5/20	7 p = 7/20
Not love Math	4 p = 4/20	9 p = 9/20	13 p = 13/20
Total	6 P=6/20	14 P=14/20	

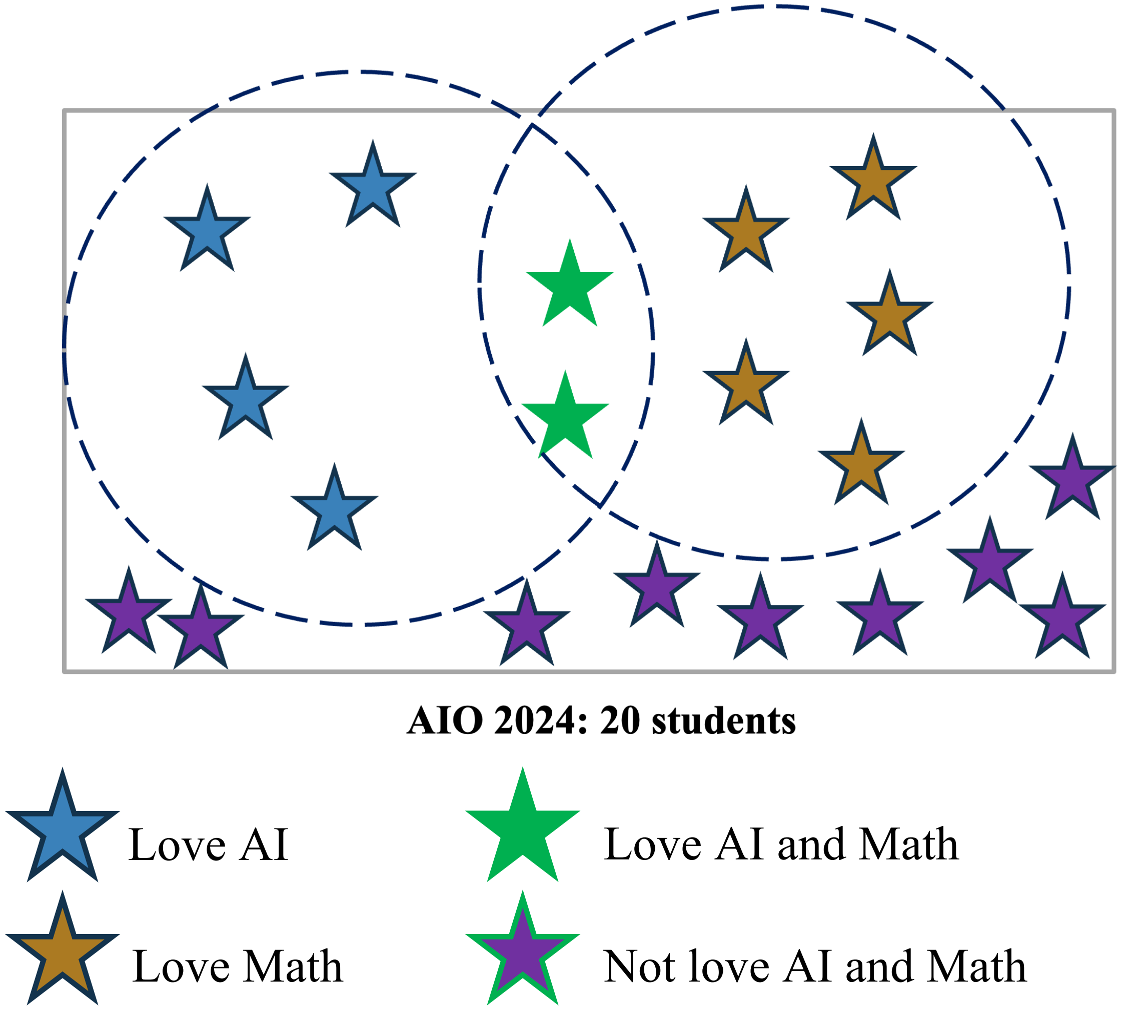
Given a student love AI (A), what is the probability that the student also love Math (B)

p(love Math | love AI) =

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{2/20}{6/20} = \frac{2}{6}$$

Review

Practice



	Love AI	Not love AI	Total
Love Math	2 p = 2/20	5 p = 5/20	7 p = 7/20
Not love Math	4 p = 4/20	9 p = 9/20	13 p = 13/20
Total	6 P=6/20	14 P=14/20	

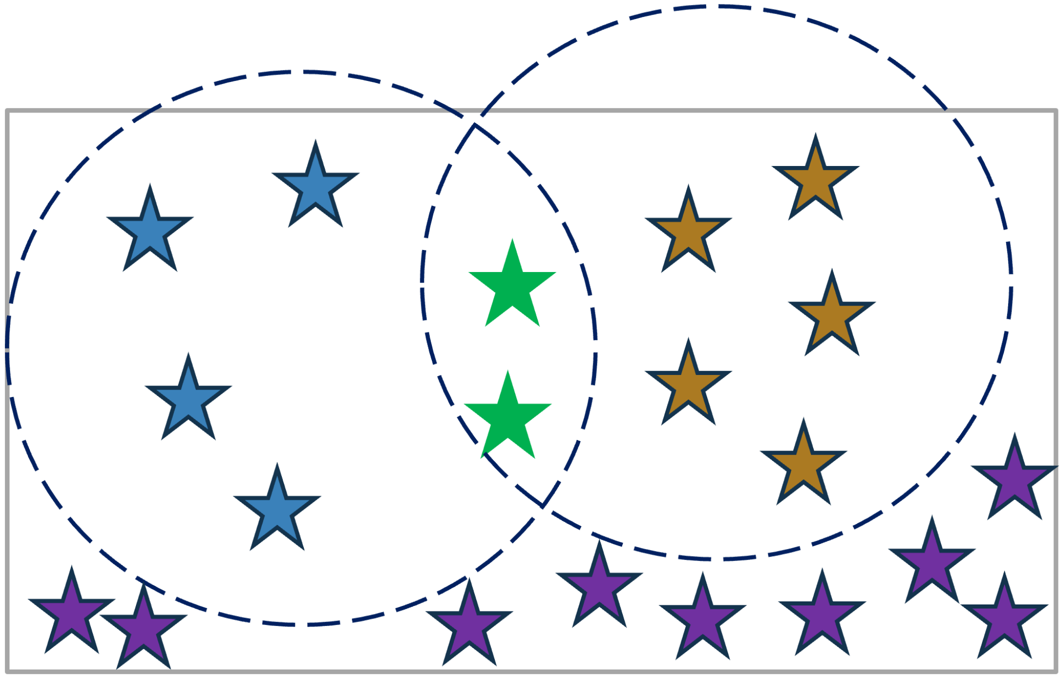
Given a student love Math (A), what is the probability that the student also love AI (B)

p(love AI | love Math) =

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{2/20}{7/20} = \frac{2}{7}$$

Review

Practice



AIO 2024: 20 students

	Love AI	Not love AI	Total
Love Math	2 p = 2/20	5 p = 5/20	7 p = 7/20
Not love Math	4 p = 4/20	9 p = 9/20	13 p = 13/20
Total	6 P=6/20	14 P=14/20	

Given a student love Math (A), what is the probability that the student not love AI (B)

p(not love AI | love Math) =

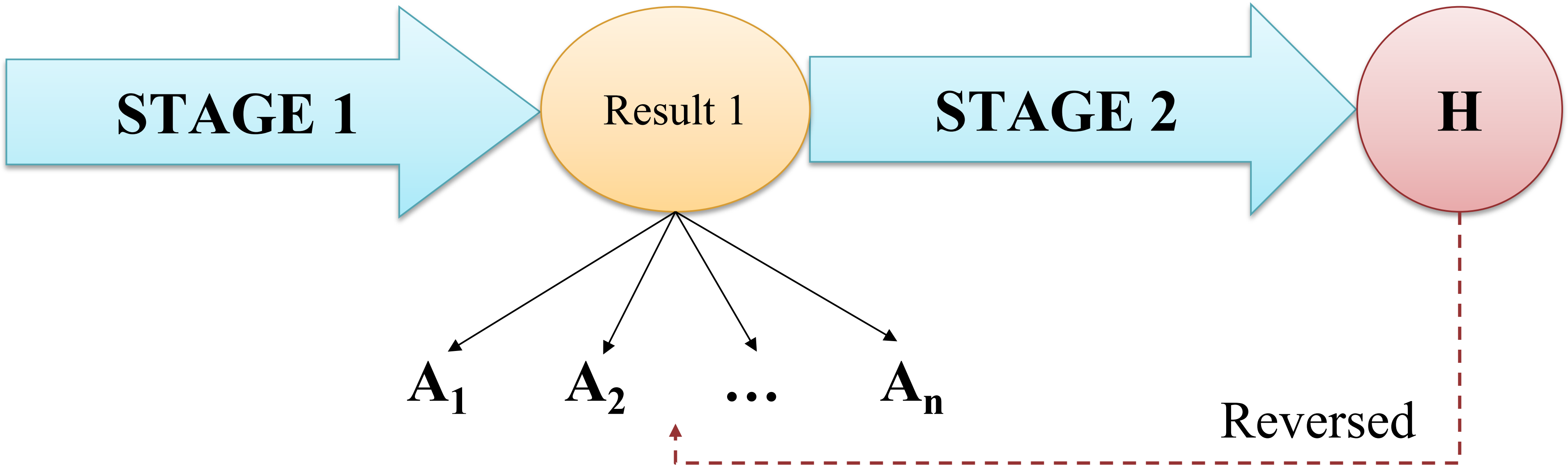
$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{5/20}{7/20} = \frac{5}{7}$$

Review

Bayes' Rule

- If $A_1, A_2, \ldots A_n$: complete system of events and H is any event with $P(A) \neq 0$:

$$P(A_i|H) = \frac{P(A_i)P(H|A_i)}{P(H)} = \frac{P(A_i)P(H|A_i)}{\sum_{j=1}^{n} P(A_j)P(H|A_j)}, i = 1, 2, \ldots, n$$



Review

Bayes' Rule

- A light bulb factory has a good bulb rate of 90%. Before being released to the market, each bulb is quality tested. Since the test is not perfect, a good bulb with probability 0.9 is recognized as good, a failed bulb with probability 0.95 being rejected.

a) Calculate the probability that the bulb passes the quality test.

b) Calculate the probability that a failed bulb passes the quality test.

Review

Bayes’ Rule

- A light bulb factory has a good bulb rate of 90%. Before being released to the market, each bulb is quality tested. Since the test is not perfect, a good bulb with probability 0.9 is recognized as good, a failed bulb with probability 0.95 being rejected.

a) Calculate the probability that the bulb passes the quality test.

b) Calculate the probability that a failed bulb passes the quality test.

1000 light bulbs

- 1000*0.9=900 → Good light bulb
  - Pass: 900*0.9=810
  - Reject: 900*0.1=90
- 1000*0.1=100 → Fail light bulb
  - Pass: 100*0.05=5
  - Reject: 100*0.95=95

Review

Bayes' Rule

- Solution:

  Let A_1: "Good light bulb", A_2: "Fail light bulb" : complete system of events

  $\Rightarrow P(A_1) = 0.9; P(A_2) = 0.1$

  H: "The light bulb passes the quality test"

  $\Rightarrow P(H|A_1) = 0.9; P(H|A_2) = 0.05$

a) The probability that the bulb passes the quality test:

  $\Rightarrow P(H) = P(A_1).P(H|A_1) + P(A_2).P(H|A_2) = 0.9*0.9 + 0.1*0.05 = 0.815$

a) The probability that a failed bulb passes the quality test:

  $\Rightarrow P(A2|H) = \frac{P(A_2)P(H|A_2)}{P(H)} = (0.1*0.05)/(0.815) = 0.0061$

Outline

SECTION 1
Review

SECTION 2
Naïve Bayes Classifier

SECTION 3
Exercise

SECTION 4
Implementation

Naïve Bayes Classifier

Classification Problem

- Input:

  A fixed set of classes $C = \{c_1, c_2,\dots, c_L\}$

  A training set of M samples: $S = \{(X_1, c_1), (X_2, c_2), (X_3, c_1),\dots (X_M, c_j)\}$

  $X = <x_1, x_2,\dots x_N>$

- Output:

  Predict a sample: $X' \Rightarrow \{c_1, c_2,\dots, c_L\}$?

Naïve Bayes Classifier

Classification Problem

- The classification problem may be formalized using a-posterior probabilities:

  $P(c|X)$ = probability that the sample $X = <x_1, x_2, \ldots, x_N>$ is of class c

  P(Result="Fail"|Confident="Yes", Studied="Yes", Sick="Yes")

  P(Result="Pass"|Confident="Yes", Studied="Yes", Sick="Yes")

  => Idea: assign to sample X the class label c such that $P(c|X)$ is maximal

Naïve Bayes Classifier

Bayes' Rule

LIKELIHOOD
The probability of "X" being True. Given "c" True

PRIOR
The probability of "c" being True. This is the knowledge

$$\mathbf{P(c|X)} = \frac{\mathbf{P(X|c).P(c)}}{\mathbf{P(X)}}$$

POSTERIOR
The probability of "c" being True. Given "X" True

MARGINALIZATION
The probability of "X" being True.

Naïve Bayes Classifier

Maximum A Posterior

$$\theta = \underset{\theta}{\text{argmax}}\, \underbrace{p(\theta|x_1, x_2, \ldots, x_N)}_{\text{posterior}}$$

$$\theta = \underset{\theta}{\text{argmax}}\, p(\theta|x_1, x_2, \ldots, x_N) = \underset{\theta}{\text{argmax}} \left[\frac{\overbrace{p(x_1, x_2, \ldots, x_N|\theta)}^{likelihood}\ \overbrace{p(\theta)}^{prior}}{\underbrace{p(x_1, x_2, \ldots, x_N)}_{evidence}} \right]$$

Independent of θ

$$\theta = \underset{\theta}{\text{argmax}}[p(x_1, x_2, \ldots, x_N|\theta)p(\theta)]$$

Naïve Bayes Classifier

Naive Bayes Classification

$$\mathbf{P(c|X) = \frac{P(X|c).P(c)}{P(X)}} \xrightarrow{\text{MAP}} \mathbf{P(c|X) \propto P(X|c).P(c) = P(x_1, x_2, \dots, x_N|c).P(c)}$$

- Maximum Likelihood Estimation (MLE)

  Assumption: all input feature are conditionally independent!

$$P(x_1, x_2, \dots, x_N|c) = P(x_1|c).P(x_2|c) \dots P(x_N|c)$$

$$\mathbf{P(c|X) \propto P(X|c).P(c) = P(\mathit{x}_1|\mathit{c}).P(\mathit{x}_2|\mathit{c}) \dots P(\mathit{x}_N|\mathit{c}).P(c)}$$

Naïve Bayes Classifier

Discrete-Valued Features Algorithm

- Training Phase: Given a training set S (M sample)

  For each target value of c (c in C)

  P(c) with examples in S

  For every feature value in x_{ij} of each feature X_i(i=1,…,M; i=1,…,N)

  compute $P(x_{ij}|c)$ with examples in S

  Output: conditional probability tables

- Test Phase: Given unknown instance X'=(x'$_1$,…,x'$_N$)

  For c in C:

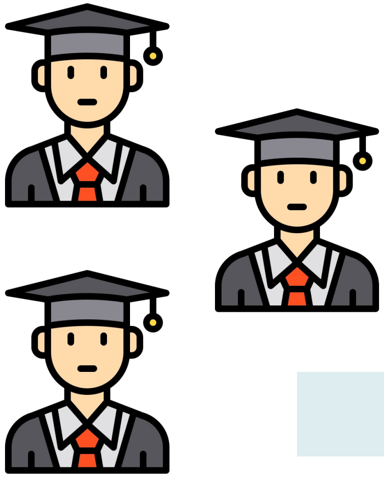
  Compute: $\mathrm{P(c|X)} \propto \mathrm{P(X|c).P(c)} = P(x_1|c).P(x_2|c)\ldots P(x_N|c).\mathrm{P(c)}$

  Choice best c

Naïve Bayes Classifier

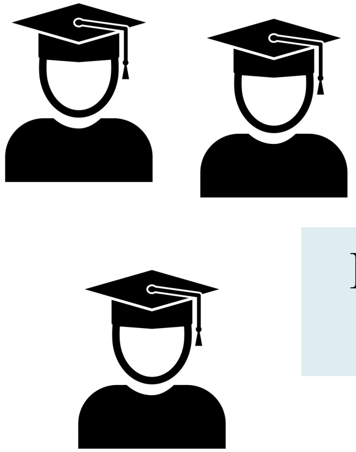
Naive Bayes Classifier for Continuous Data

Math	Art	English
9.5	7.5	5.5
8.2	8.0	6.5
7.0	9.0	7.0
…	…	…



Love AI

Math	Art	English
1.5	6.5	8.5
5.0	8.5	8.5
9.0	8.0	8.0
…	…	…



Does not love AI

AI VIET NAM
@aivietnam.edu.vn

Naïve Bayes Classifier

Naive Bayes Classifier for Continuous Data

	Math	Art	English
	9.5	7.5	5.5
	8.2	8.0	6.5
	7.0	9.0	7.0
Mean (μ)			
Std (σ)			

$$\mu = \frac{\sum x}{n}$$

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

	Math	Art	English
	1.5	6.5	8.5
	5.0	8.5	8.5
	9.0	8.0	8.0
Mean (μ)	…	…	…
Std(σ)			

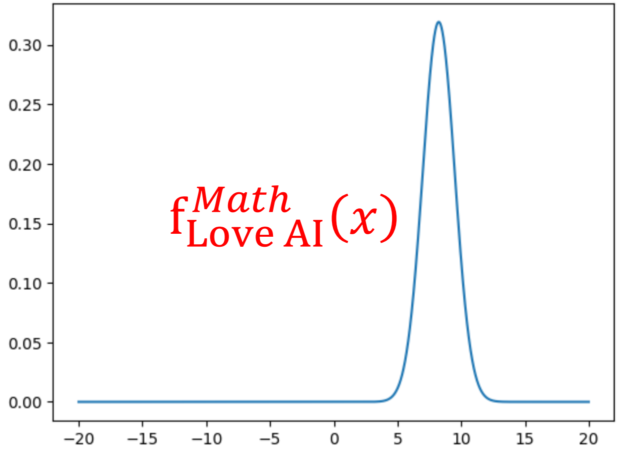
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Naïve Bayes Classifier

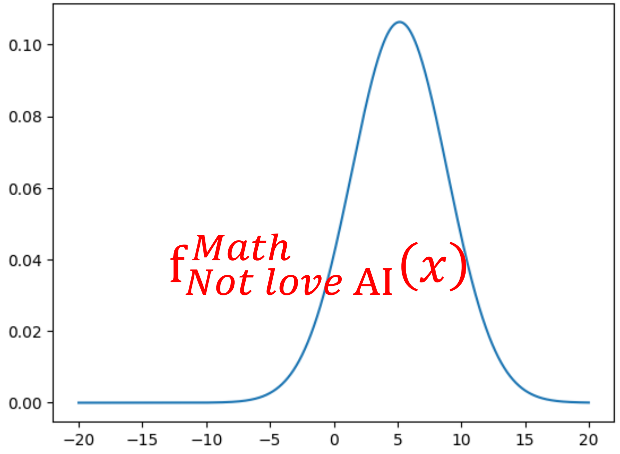
Naive Bayes Classifier for Continuous Data

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

	Math	Art	English
	9.5	7.5	5.5
	8.2	8.0	6.5
	7.0	9.0	7.0
Mean (μ)			
Std (σ)			



	Math	Art	English
	1.5	6.5	8.5
	5.0	8.5	8.5
	9.0	8.0	8.0
Mean (μ)			
Std(σ)			



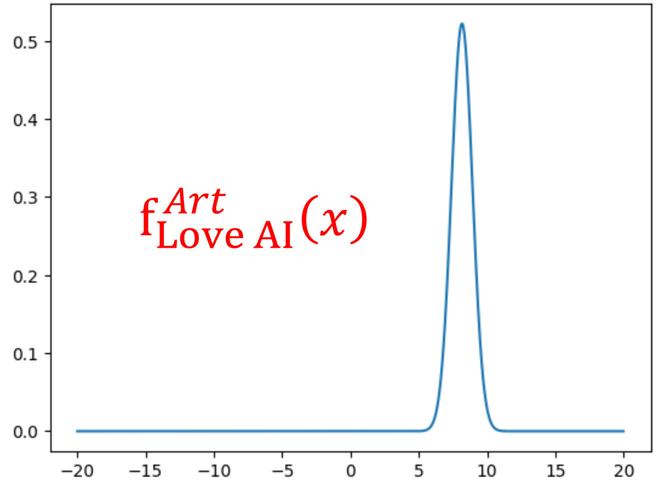
AI VIET NAM
@aivietnam.edu.vn

Naïve Bayes Classifier

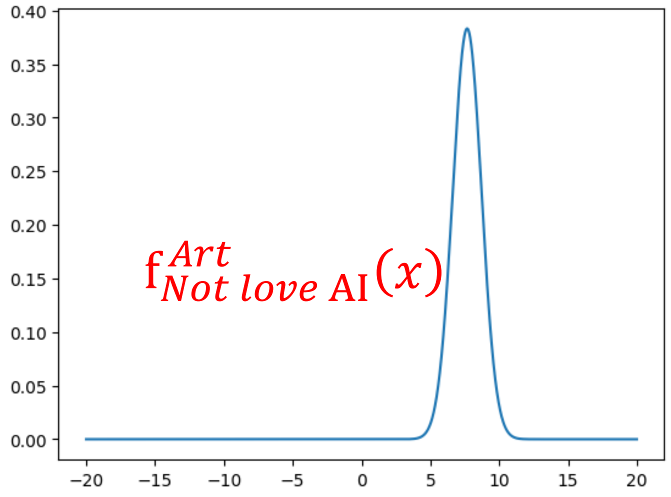
Naive Bayes Classifier for Continuous Data

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

	Math	Art	English
	9.5	7.5	5.5
	8.2	8.0	6.5
	7.0	9.0	7.0
Mean (μ)			
Std (σ)			



	Math	Art	English
	1.5	6.5	8.5
	5.0	8.5	8.5
	9.0	8.0	8.0
Mean (μ)			
Std(σ)			

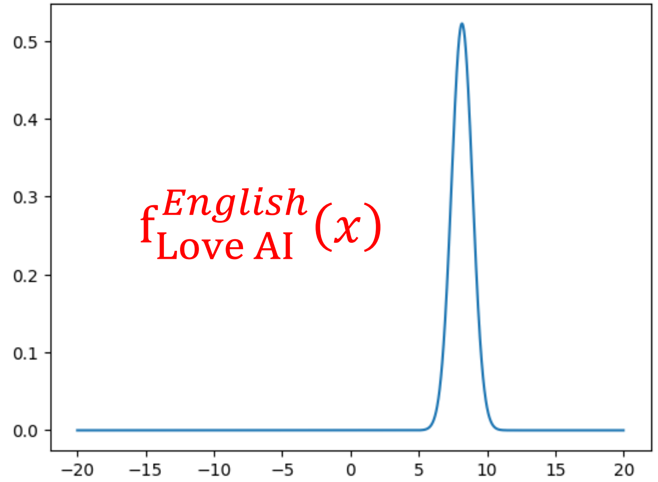


Naïve Bayes Classifier

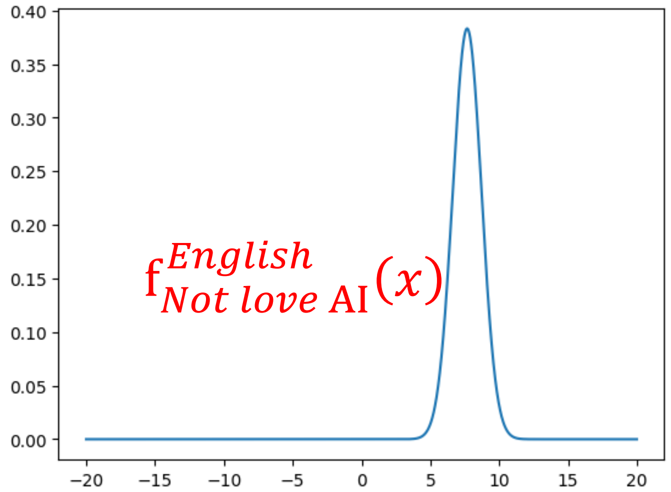
Naive Bayes Classifier for Continuous Data

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

	Math	Art	English
	9.5	7.5	5.5
	8.2	8.0	6.5
	7.0	9.0	7.0
Mean (μ)			
Std (σ)			



	Math	Art	English
	1.5	6.5	8.5
	5.0	8.5	8.5
	9.0	8.0	8.0
Mean (μ)			
Std(σ)			



Naïve Bayes Classifier

Naive Bayes Classifier for Continuous Data

Does she love AI or Not?

Given a student information (math, art, and english scores), what is the probability that he/she loves AI or not.
(Math = 5 & Art = 6 & English = 7)

Naïve Bayes Classifier

Naive Bayes Classifier for Continuous Data

Does she love AI or Not?

P (Love AI |Math = 5 & Art = 6 & English = 7)
= P (Math = 5 & Art = 6 & English = 7 | Love AI) . P(Love AI)
= P (Math = 5| Love AI) . P(Art = 6| Love AI) . P(English = 7 | Love AI) . P(Love AI)

$\mathrm{f}_{\text{Love AI}}^{Math}(x=5)$ $\mathrm{f}_{love\ \text{AI}}^{Art}(x=6)$ $\mathrm{f}_{love\ \text{AI}}^{English}(x)=7$

P (Not Love AI |Math = 5 & Art = 6 & English = 7)
= P (Math = 5 & Art = 6 & English = 7 | Not Love AI) . P(Not Love AI)
= P (Math = 5| Not Love AI) . P(Art = 6| Not Love AI) . P(English = 7 | Not Love AI) . P(Not Love AI)

$\mathrm{f}_{not\ \text{Love AI}}^{Math}(x=5)$ $\mathrm{f}_{not\ love\ \text{AI}}^{Art}(x=6)$ $\mathrm{f}_{not\ love\ \text{AI}}^{English}(x)=7$

Outline

SECTION 1
Review

SECTION 2
Naïve Bayes Classifier

SECTION 3
Exercise

SECTION 4
Implementation

Exercise

Exercise 1: PLAY TENNIS

- Training Samples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Overcast	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Exercise

Exercise 1: PLAY TENNIS

- Training Samples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Overcast	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Exercise

Exercise 1: PLAY TENNIS

➢ Probability Tables

Day	Outlook	Temp	Hum	Wind	PlayT
D1	Sunny	Hot	High	Weak	No
	Sunny	Hot	High	Strong	No
	Overcast	Hot	High	Weak	Yes
	Rain	Mild	High	Weak	Yes
	Rain	Cool	Normal	Weak	Yes
	Rain	Cool	Normal	Strong	No
	Overcast	Cool	Normal	Strong	Yes
	Overcast	Mild	High	Weak	No
	Sunny	Cool	Normal	Weak	Yes
	Rain	Mild	Normal	Weak	Yes

Attribute		Class	
		Yes	No
Prior Probability		6/10	4/10
Outlook	Sunny		
	Overcast		
	Rain		
Temperature	Hot		
	Mild		
	Cool		
Humidity	High		
	Normal		
Wind	Weak		
	Strong		

Exercise

Exercise 1: PLAY TENNIS

- Probability Tables

Day	Outlook	Temp	Hum	Wind	PlayT
D1	Sunny	Hot	High	Weak	No
	Sunny	Hot	High	Strong	No
	Overcast	Hot	High	Weak	Yes
	Rain	Mild	High	Weak	Yes
	Rain	Cool	Normal	Weak	Yes
	Rain	Cool	Normal	Strong	No
	Overcast	Cool	Normal	Strong	Yes
	Overcast	Mild	High	Weak	No
	Sunny	Cool	Normal	Weak	Yes
	Rain	Mild	Normal	Weak	Yes

Attribute		Class	
		Yes	No
Prior Probability		6/10	4/10
Outlook	Sunny	1/6	2/4
	Overcast	2/6	1/4
	Rain	3/6	1/4
Temperature	Hot	1/6	2/4
	Mild	2/6	1/4
	Cool	3/6	1/4
Humidity	High	2/6	3/4
	Normal	4/6	1/4
Wind	Weak	5/6	2/4
	Strong	1/6	2/4

Exercise

Exercise 1: PLAY TENNIS

- Test Phase

D11	Sunny	Cool	High	Strong	?

Attribute		Class	
		Yes	No
Prior Probability		6/10	4/10
Outlook	Sunny	1/6	2/4
	Overcast	2/6	1/4
	Rain	3/6	1/4
Temperature	Hot	1/6	2/4
	Mild	2/6	1/4
	Cool	3/6	1/4
Humidity	High	2/6	3/4
	Normal	4/6	1/4
Wind	Weak	5/6	2/4
	Strong	1/6	2/4

P("Play Tennis"="Yes"|X)

P(X|"Play Tennis"="Yes").P("Play Tennis"="Yes")
=P("Outlook"="Sunny"|"Play Tennis"="Yes")
. P("Temp"="Cool"|"Play Tennis"="Yes")
. P("Hum"="High"|"Play Tennis"="Yes")
. P("Wind"="Strong"|"Play Tennis"="Yes")
. P("Play Tennis"="Yes")
=1/6.3/6.2/6.1/6.6/10
=0.0028

Exercise

Exercise 1: PLAY TENNIS

- Test Phase

Attribute		Class	
		Yes	No
Prior Probability		6/10	4/10
Outlook	Sunny	1/6	2/4
	Overcast	2/6	1/4
	Rain	3/6	1/4
Temperature	Hot	1/6	2/4
	Mild	2/6	1/4
	Cool	3/6	1/4
Humidity	High	2/6	3/4
	Normal	4/6	1/4
Wind	Weak	5/6	2/4
	Strong	1/6	2/4

D11	Sunny	Cool	High	Strong	?

P(“Play Tennis”=“No”|X)

P(X|“Play Tennis”=“No”).P(“Play Tennis”=“No”)
=P(“Outlook”=“Sunny”|“Play Tennis”=“No”)
. P(“Temp”=“Cool”|“Play Tennis”=“No”)
. P(“Hum”=“High”|“Play Tennis”=“No”)
. P(“Wind”=“Strong”|“Play Tennis”=“No”)
. P(“Play Tennis”=“No”)
=2/4.1/4.3/4.2/4.4/10
=0.0188

Exercise

Exercise 1: PLAY TENNIS

- Test Phase

Attribute		Class	
		Yes	No
Prior Probability		6/10	4/10
Outlook	Sunny	1/6	2/4
	Overcast	2/6	1/4
	Rain	3/6	1/4
Temperature	Hot	1/6	2/4
	Mild	2/6	1/4
	Cool	3/6	1/4
Humidity	High	2/6	3/4
	Normal	4/6	1/4
Wind	Weak	5/6	2/4
	Strong	1/6	2/4

D11	Sunny	Cool	High	Strong	?

P("Play Tennis"="Yes"|X) $\propto$ 0.0028

P("Play Tennis"="No"|X) $\propto$ 0.0188

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

- Training Samples

Day	Season	Fog	Rain	Class
Weekday	Spring	None	None	On Time
Weekday	Winter	None	Slight	On Time
Weekday	Winter	None	None	On Time
Holiday	Winter	High	Slight	Late
Saturday	Summer	Normal	None	On Time
Weekday	Autumn	Normal	None	Very Late
Holiday	Summer	High	Slight	On Time
Sunday	Summer	Normal	None	On Time
Weekday	Winter	High	Heavy	Very Late
Weekday	Summer	None	Slight	On Time
Saturday	Spring	High	Heavy	Cancelled
Weekday	Summer	High	Slight	On Time
Weekday	Winter	Normal	None	Late
Weekday	Summer	High	None	On Time
Weekday	Winter	Normal	Heavy	Very Late
Saturday	Autumn	High	Slight	On Time
Weekday	Autumn	None	Heavy	On Time
Holiday	Spring	Normal	Slight	On Time
Weekday	Spring	Normal	None	On Time
Weekday	Spring	Normal	Heavy	On Time

AI VIET NAM
@aivietnam.edu.vn

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

Day	Season	Fog	Rain	Class
Weekday	Spring	None	None	On Time
Weekday	Winter	None	Slight	On Time
Weekday	Winter	None	None	On Time
Holiday	Winter	High	Slight	Late
Saturday	Summer	Normal	None	On Time
Weekday	Autumn	Normal	None	Very Late
Holiday	Summer	High	Slight	On Time
Sunday	Summer	Normal	None	On Time
Weekday	Winter	High	Heavy	Very Late
Weekday	Summer	None	Slight	On Time
Saturday	Spring	High	Heavy	Cancelled
Weekday	Summer	High	Slight	On Time
Weekday	Winter	Normal	None	Late
Weekday	Summer	High	None	On Time
Weekday	Winter	Normal	Heavy	Very Late
Saturday	Autumn	High	Slight	On Time
Weekday	Autumn	None	Heavy	On Time
Holiday	Spring	Normal	Slight	On Time
Weekday	Spring	Normal	None	On Time
Weekday	Spring	Normal	Heavy	On Time

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability					
Day	Weekday				
	Holiday				
	Sunday				
	Saturday				
Season	Spring				
	Winter				
	Summer				
	Autumn				
Fog	None				
	High				
	Normal				
Rain	None				
	Slight				
	Heavy				

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

Day	Season	Fog	Rain	Class
Weekday	Spring	None	None	On Time
Weekday	Winter	None	Slight	On Time
Weekday	Winter	None	None	On Time
Holiday	Winter	High	Slight	Late
Saturday	Summer	Normal	None	On Time
Weekday	Autumn	Normal	None	Very Late
Holiday	Summer	High	Slight	On Time
Sunday	Summer	Normal	None	On Time
Weekday	Winter	High	Heavy	Very Late
Weekday	Summer	None	Slight	On Time
Saturday	Spring	High	Heavy	Cancelled
Weekday	Summer	High	Slight	On Time
Weekday	Winter	Normal	None	Late
Weekday	Summer	High	None	On Time
Weekday	Winter	Normal	Heavy	Very Late
Saturday	Autumn	High	Slight	On Time
Weekday	Autumn	None	Heavy	On Time
Holiday	Spring	Normal	Slight	On Time
Weekday	Spring	Normal	None	On Time
Weekday	Spring	Normal	Heavy	On Time

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

AI VIET NAM
@aivietnam.edu.vn

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

➢ Test Phase

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

Day	Season	Fog	Rain	Class
Weekday	Winter	High	Heavy	?

P(“Class”=“On Time”|X)

P(X|“Class”=“On Time”)P(“Class”=“On Time”
=P(“Day”=“Weekday”|Class”=“On Time”)
. P(“Season”=“Winter”|“Class”=“On Time”)
. P(“Fog”=“High”|“Class”=“On Time”)
. P(“Rain”=“Heavy”|“Class”=“On Time”)
. P(“Class”=“One Time”)
=9/14.2/14.4/14.2/14.14/20
=0.0026

AI VIET NAM
@aivietnam.edu.vn

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

- Test Phase

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

Day	Season	Fog	Rain	Class
Weekday	Winter	High	Heavy	?

P(“Class”=“Late”|X)

P(X|“Class”=“Late”)P(“Class”=“Late”
=P(“Day”=“Weekday”|Class”=“Late”)
. P(“Season”=“Winter”|“Class”=“Late”)
. P(“Fog”=“High”|“Class”=“Late”)
. P(“Rain”=“Heavy”|“Class”=“Late”)
. P(“Class”=“Late”)
=1/2.2/2.1/2.0/2.2/20
=0.0000

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

- Test Phase

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

Day	Season	Fog	Rain	Class
Weekday	Winter	High	Heavy	?

P("Class"="Very Late"|X)

P(X|"Class"="Very Late")P("Class"="Very Late"
=P("Day"="Weekday"|Class"="Very Late")
. P("Season"="Winter"|"Class"="Very Late")
. P("Fog"="High"|"Class"="Very Late")
. P("Rain"="Heavy"|"Class"="Very Late")
. P("Class"="Very Late")
=3/3.2/3.1/3.2/3.3/20
=0.0222

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

- Test Phase

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

Day	Season	Fog	Rain	Class
Weekday	Winter	High	Heavy	?

P("Class"="Cancelled"|X)

P(X|"Class"="Cancelled")P("Class"="Cancelled"
=P("Day"="Weekday"|Class"="Cancelled")
. P("Season"="Winter"|"Class"="Cancelled")
. P("Fog"="High"|"Class"="Cancelled")
. P("Rain"="Heavy"|"Class"="Cancelled")
. P("Class"="Cancelled")
=0/1.0/1.1/1.1/1.1/20
=0.0000

AI VIET NAM
@aivietnam.edu.vn

Exercise

Exercise 2: TRAFFIC DATA (MULTI-LABEL CLASSIFICATION)

- Test Phase

Attribute		Class			
		On Time	Late	Very Late	Cancelled
Prior Probability		14/20	2/20	3/20	1/20
Day	Weekday	9/14	1/2	3/3	0/1
	Holiday	2/14	1/2	0/3	0/1
	Sunday	1/14	0/2	0/3	0/1
	Saturday	2/14	0/2	0/3	1/1
Season	Spring	4/14	0/2	0/3	1/1
	Winter	2/14	2/2	2/3	0/1
	Summer	6/14	0/2	0/3	0/1
	Autumn	2/14	0/2	1/3	0/1
Fog	None	5/14	0/2	0/3	0/1
	High	4/14	1/2	1/3	1/1
	Normal	5/14	1/2	2/3	0/1
Rain	None	6/14	1/2	1/3	0/1
	Slight	6/14	1/2	0/3	0/1
	Heavy	2/14	0/2	2/3	1/1

Day	Season	Fog	Rain	Class
Weekday	Winter	High	Heavy	?

P(“Class”=“On Time”|X) ∝ 0.0026

P(“Class”=“Late”|X) ∝ 0.0000

P(“Class”=“Very Late”|X) ∝ 0.0222

P(“Class”=“Cancelled”|X) ∝ 0.0000

QUIZ TIME

Exercise

Exercise 3: IRIS CLASSIFICATION

- Training Phase

Length	1.4	1.0	1.3	1.9	2.0	1.8	3.0	3.8	4.1	3.9	4.2	3.4
Class	0	0	0	0	0	0	1	1	1	1	1	1

$$\mu = \frac{\sum x}{n}$$

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

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

$$\mu = \frac{\sum x}{n} = \frac{1.4 + 1.0 + 1.3 + 1.9 + 2.0 + 1.8}{6}$$

Class	Mean	Variance
0	1.56	0.128
1		

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

$$= \frac{(1.4 - 1.56)^2 + (1.0 - 1.56)^2 + (1.3 - 1.56)^2 + (1.9 - 1.56)^2 + (2.0 - 1.56)^2 + (1.8 - 1.56)^2}{6}$$

Exercise

Exercise 3: IRIS CLASSIFICATION

- Training Phase

Length	1.4	1.0	1.3	1.9	2.0	1.8	3.0	3.8	4.1	3.9	4.2	3.4
Class	0	0	0	0	0	0	1	1	1	1	1	1

$$\mu = \frac{\sum x}{n}$$

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

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

$$\mu = \frac{\sum x}{n} = \frac{3.7 + 3.8 + 4.1 + 3.9 + 4.2 + 3.4}{6}$$

Class	Mean	Variance
0	1.56	0.128
1	3.73	0.172

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

$$= \frac{(3.7 - 3.73)^2 + (3.8 - 3.73)^2 + (4.1 - 3.73)^2 + (3.9 - 3.73)^2 + (4.2 - 3.73)^2 + (3.4 - 3.73)^2}{6}$$

Exercise

Exercise 3: IRIS CLASSIFICATION

- Test Phase (Length X=3.4)

Length	1.4	1.0	1.3	1.9	2.0	1.8	3.0	3.8	4.1	3.9	4.2	3.4
Class	0	0	0	0	0	0	1	1	1	1	1	1

$$\mu = \frac{\sum x}{n}$$

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

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

P("Class"="0"|X) ∝ P(X|"Class"="0").P("Class"="0")

P(X|"Class"="0")= $\frac{1}{\sqrt{2\pi * 0.128}} e^{-\frac{1}{2}\left(\frac{3.4-1.56}{\sqrt{0.128}}\right)^2} = 2.18 * 10^{-6}$

Class	Mean	Variance
0	1.56	0.128
1	3.73	0.172

P("Class"="0"|X) ∝ P(X|"Class"="0").P("Class"="0")= $2.18 * 10^{-6} * 6/12$= 1.09*10^{-6}

Exercise

Exercise 3: IRIS CLASSIFICATION

- Test Phase (Length X=3.4)

Length	1.4	1.0	1.3	1.9	2.0	1.8	3.0	3.8	4.1	3.9	4.2	3.4
Class	0	0	0	0	0	0	1	1	1	1	1	1

$$\mu = \frac{\sum x}{n}$$

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

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

P("Class"="1"|X) $\propto$ P(X|"Class"="1").P("Class"="1")

P(X|"Class"="1")= $\frac{1}{\sqrt{2\pi * 0.172}} e^{-\frac{1}{2}\left(\frac{3.4-3.73}{\sqrt{0.172}}\right)^2} = 0.697$

Class	Mean	Variance
0	1.56	0.128
1	3.73	0.172

P("Class"="1"|X) $\propto$ P(X|"Class"="1").P("Class"="1")= $0.697 * 6/12$= 0.3486

Exercise

Exercise 3: IRIS CLASSIFICATION

- Test Phase (Length X=3.4)

Length	1.4	1.0	1.3	1.9	2.0	1.8	3.0	3.8	4.1	3.9	4.2	3.4
Class	0	0	0	0	0	0	1	1	1	1	1	1

$$\mu = \frac{\sum x}{n}$$

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

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

P("Class"="0"|X)
$\propto$ P(X|"Class"="0").P("Class"="0")
$= 2.18 * 10^{-6} * 6/12 = 1.09*10^{-6}$

Class	Mean	Variance
0	1.56	0.128
1	3.73	0.172

P("Class"="1"|X)
$\propto$ P(X|"Class"="1").P("Class"="1")
$= 0.697 * 6/12 = 0.3486$

Outline

SECTION 1
Review

SECTION 3
Exercise

SECTION 2
Naïve Bayes Classifier

SECTION 4
Implementation

Implementation

PLAY TENNIS CLASSIFIER

Cho trước dữ liệu thời tiết của 10 ngày (D1-D10, như bảng 1). Hãy phát triển chương trình sử dụng mô hình phân loại Naive Bayes để dự đoán xem ngày thứ 11 (D11), AD có thể chơi tennis hay không?

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D11	Sunny	Cool	High	Strong	???

(a) "Play Tennis" = "Yes"

(b) "Play Tennis" = "No"

Implementation

IRIS CLASSIFIER

Cho trước dữ liệu chứa thông tin về hoa Iris gồm có sepal length, sepal width và petal lenght, và Species (bảng 6). Hãy phát triển chương trình sử dụng mô hình phân loại Gausian Naive Bayes để dự đoán chủng loại của hoa Iris. Dữ liệu hoa iris được lưu trữ trong file iris_data.txt có thể được tải vềtại đây.

No.	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	6.4	3.1	5.5	1.8	Iris-virginica
4	6.0	3.0	4.8	1.8	Iris-virginica
5	6.0	2.2	4.0	1.0	Iris-versicolora
...	...	...	...	..	...

AI VIET NAM
@aivietnam.edu.vn

Summary

PROBABILITY

- Classical Probability
- Geometric Probability
- Rules of Probability
  - Addition
  - Conditional Probability
  - Multiplication
- Bayes' Rule

Naïve Bayes Classifier

- Bayes Classifier

$$P(c|X) \propto P(X|c).P(c)$$
$$= P(x_1|c).P(x_2|c)\ldots P(x_N|c).P(c)$$

$$\mu = \frac{\sum x}{n}$$

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

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

Thanks!

Any questions?