AI VIETNAM
All-in-One Course

Naïve Bayes Classifiers

TA Team
AI VIETNAM

Quang-Vinh Dinh
Ph.D. in Computer Science

Year 2022

Outline

- Basic Concepts
- Bayes' Theorem
- Bernoulli Naïve Bayes Classifier
- Gaussian Naïve Bayes Classifier
- Histogram Equalization

Basic Probability

❖ **Some concepts**

Toss a coin

Sample space: S = {heads, tails}

Experiment: implementation of set of basic conditions for observing a certain phenomenon

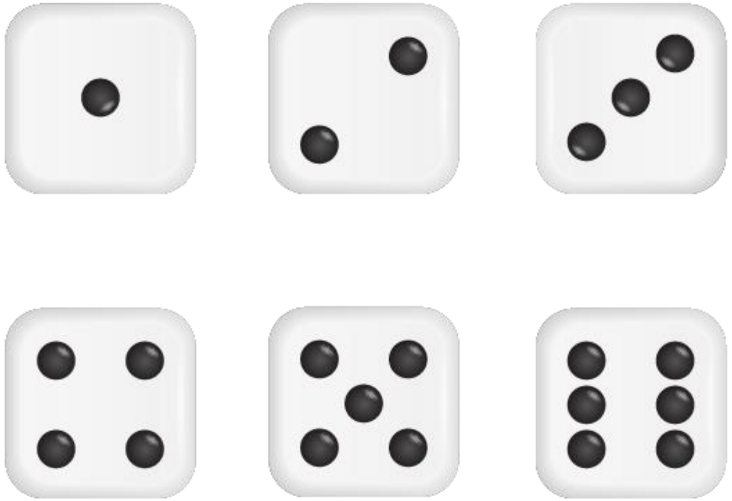
An outcome is a result of an experiment

The set of all possible outcomes is called the sample space

An event is a subset of the sample space

Roll a dice

Sample space: S = {1, 2, 3, 4, 5, 6}



❖ Event

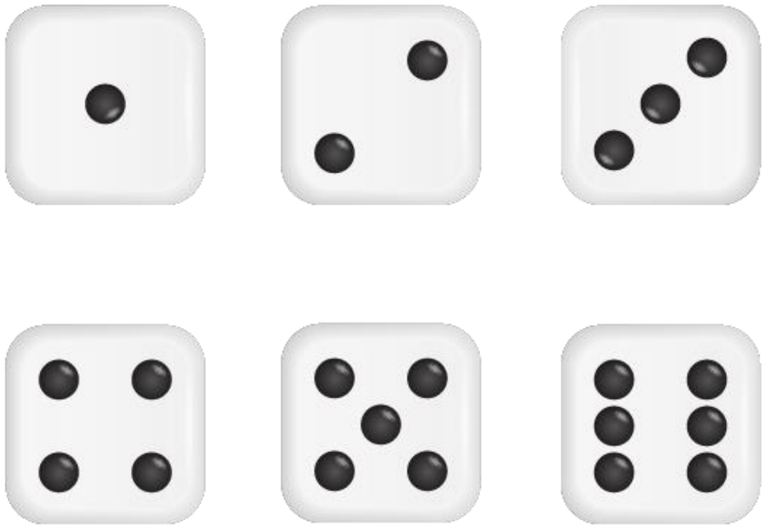
Certain event: An event that always occurs in an experiment, denoted by Ω

Impossible event: An event that never occurs when the experiment is executed, denoted by $\emptyset$.

Random event: An event that may or may not occur when performing the experiment

Random Experiment: An experiment whose outcomes are random events

For convenience, events are usually denoted with capital letters: A, B, C, . . .



Roll a dice:

Ω = “dots ≤ 6 and ≥ 1” is a certain event

$\emptyset$ = “7-dot” is an impossible event

A = “even-dot” is a random event

Experiment and Event

❖ Example

- A family with 2 children. Events:
  - A = "A family has 1 boy and 1 girl"
  - B = "A family has 3 children"
  - C = "A family has 2 children"

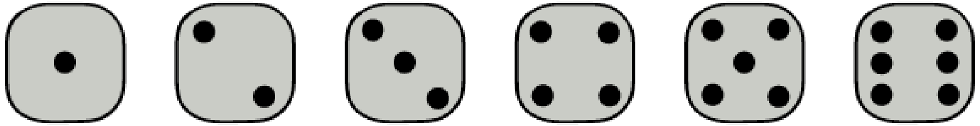
  Which event is certain random, impossible event?
- A box contains 8 balls: 6 blue and 2 red. Pick randomly 3 balls:
  - A = "get 3 blue balls"
  - B = "get 3 red balls"
  - C = "get 3 balls"

  Which event is certain random, impossible event?

Operations with Events

❖ Intersection of events

- In the experiment of rolling a single dice



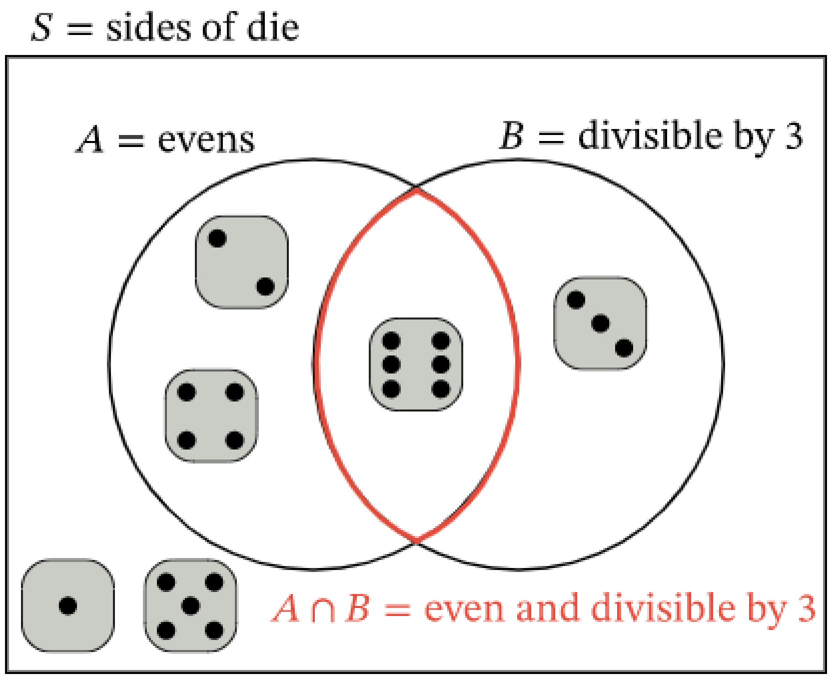
- Event A: "the number rolled is even"

  => A = {2, 4, 6}

- Event B: "the number rolled is divisible by 3"

  => B = {3, 6 }

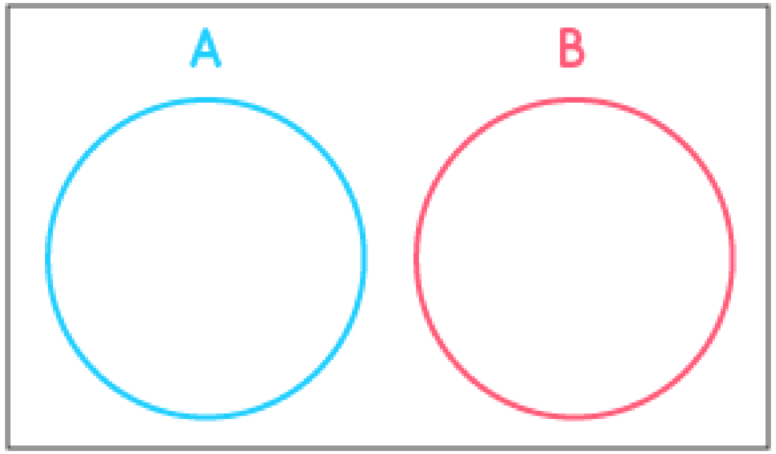
- $A \cap B = \{6\}$



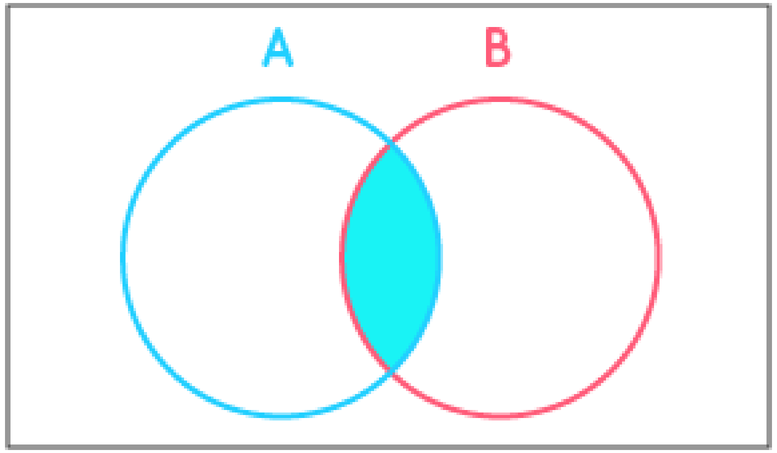
Operations with Events

❖ Mutually exclusive event

- ✓ Events A and B are mutually exclusive (cannot both occur at once) if they have no elements in common.
- ✓ For A and B to have no outcomes in common means precisely that it is impossible for both A and B to occur on a single trial of the random experiment.
- ✓ $A \cap B = \{\emptyset\}$



Mutually exclusive



Non-mutually exclusive

Example

Event A: "the number rolled is even"
=> $A = \{2, 4, 6\}$

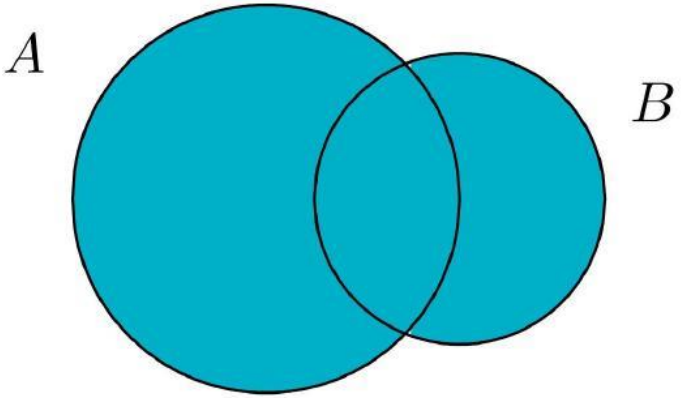
Event B: "the number rolled is odd"
=> $B = \{1, 3, 5\}$

$A \cap B = \{\emptyset\}$

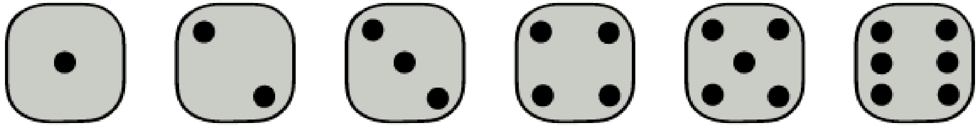
Operations with Events

- **Union of events**

  - ✓ The union of events A and B, denoted A ∪ B
  - ✓ The collection of all outcomes that are elements of one or the other of the sets A and B, or of both of them.



- In the experiment of rolling a single dice



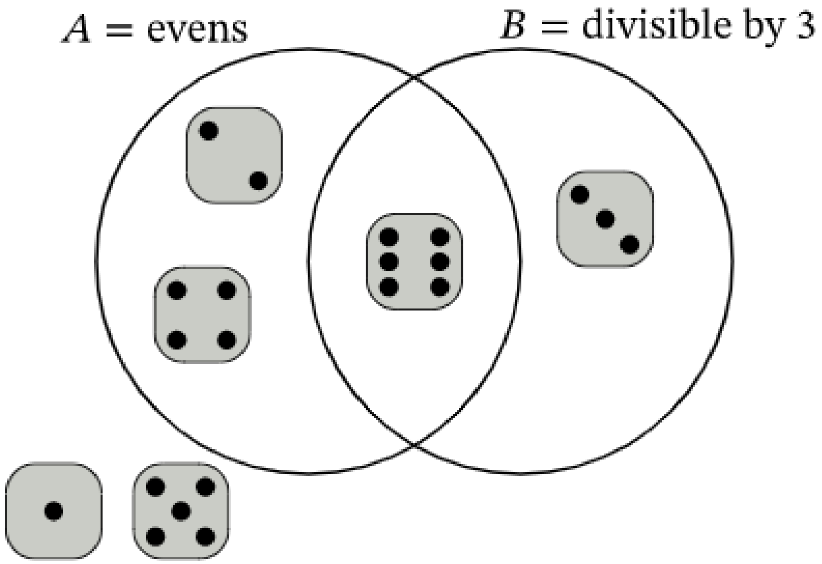
- Event A: "the number rolled is even"

  => A = {2, 4, 6}

- Event B: "the number rolled is divisible by 3"

  => B = {3, 6 }

- Find the union of A and B?

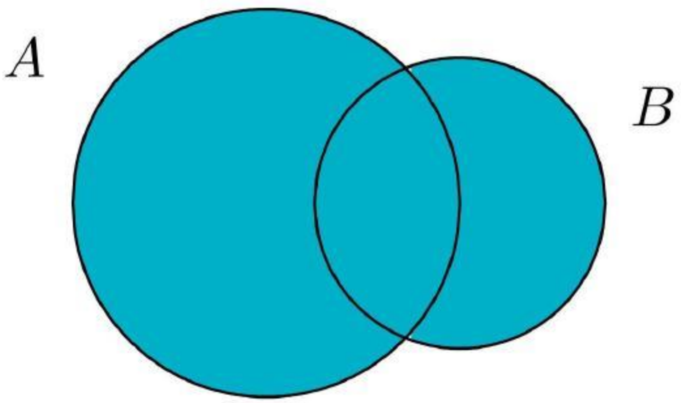


Operations with Events

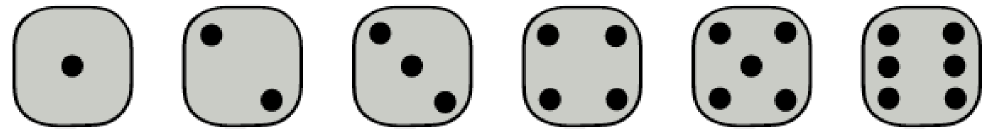
❖ **Union of events**

✓ The union of events A and B, denoted $A \cup B$

✓ The collection of all outcomes that are elements of one or the other of the sets A and B, or of both of them.



➢ In the experiment of rolling a single dice



➢ Event A: "the number rolled is even"

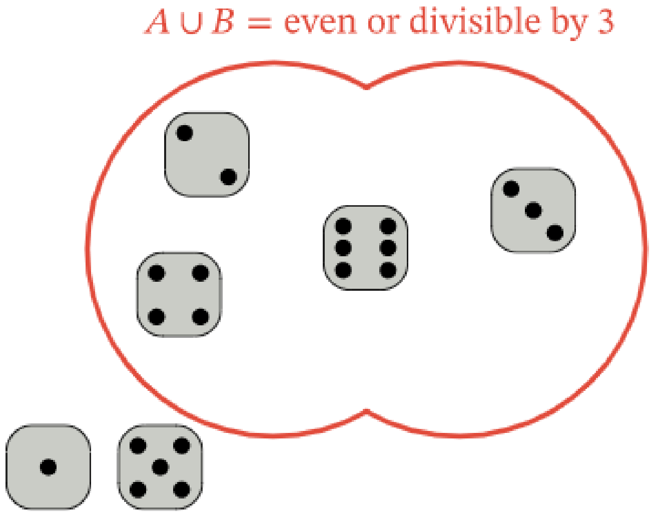
=> $A = \{2, 4, 6\}$

➢ Event B: "the number rolled is divisible by 3"

=> $B = \{3, 6\}$

➢ Find the union of A and B?

=> $A \cup B = \{2, 3, 4, 6\}$



Operations with Events

❖ Complements

✓ The complement of an event A in a sample space S, denoted A' (A^c)

✓ The collection of all outcomes in S that are not elements of the set A

✓ $A' + A = \Omega$

Example:

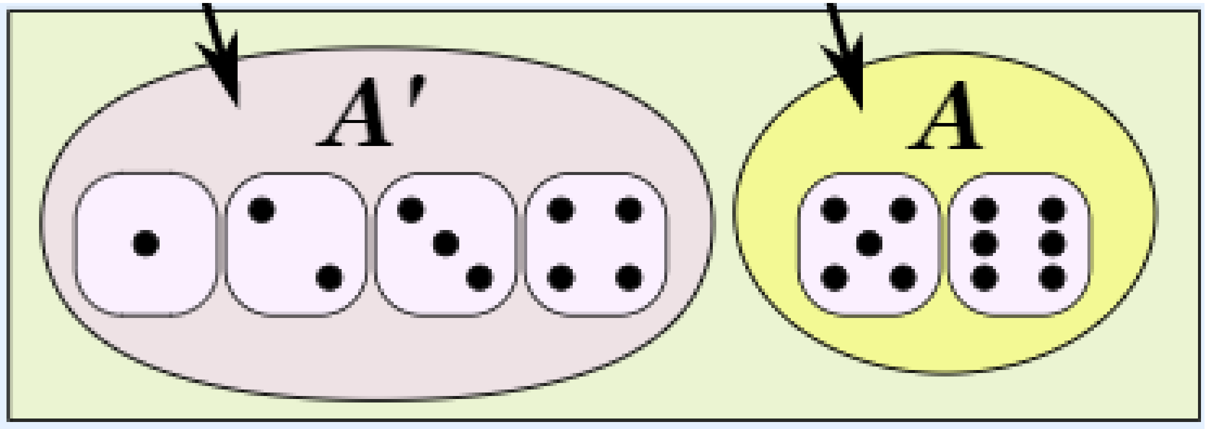
A: "the number rolled is greater than 4."

=> A = {5, 6}

=> A' = {1, 2, 3, 4}

Complement of an event A

An event A



Operations with Events

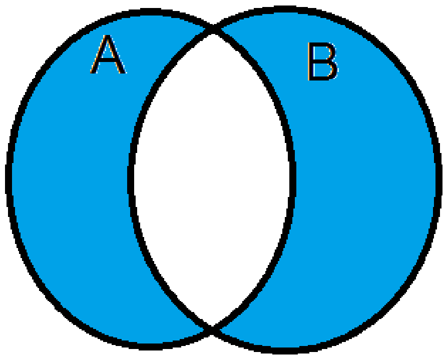
❖ Quizzes

The region colored in the figure below is represented by:

A. (A.B’).(A’.B)

B. (A + B’)(A’ + B)

C. A.B’ + A’.B



In the experiment of rolling a single dice

A: “the number rolled is even number.”

B: “the number rolled is greater than or equal 4.”

C: “the number rolled is greater than 2.”

1. Event A’ is

A. {}　　B. {1, 3, 5}　　C. {1, 3}　　D. {2, 4, 6}

2. Event A.B is

A. {5, 7}　　B. {4, 6}　　C. {5}　　D. {1, 3, 5, 6}

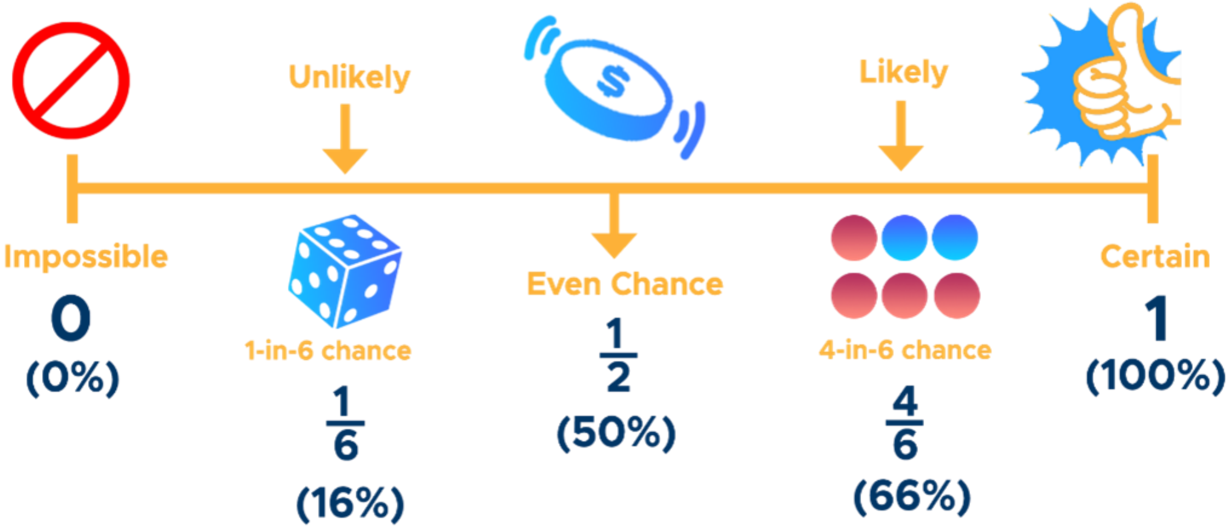
3. Event B + C is

A. {2, 4}　　B. {1, 4, 5, 6}　　C. {3, 4, 5, 6}　　D. {1, 2, 5, 6}

Probability

❖ Definition

- ✓ The probability of an event A is P(A) - a number between 0 and 1 that shows how likely the event is
- ✓ P(A) => 0: very unlikely that the event A occurs
- ✓ P(A) => 1: very likely to occur
- ✓ Some properties:
  - $0 <= P(A) <= 1$
  - $P(\Omega) = 1$
  - $P(\emptyset) = 0$



Probability

❖ Classical Probability

The theoretical probability of an event A is the number of ways the event can occur divided by the total number of possible outcomes:

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

Example

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

- There are 6 faces on a fair dice, numbered 1 to 6

=> n(Ω) = 6

- A : "even number" => A = {2, 4, 6} => n(A) = 3

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

Probability

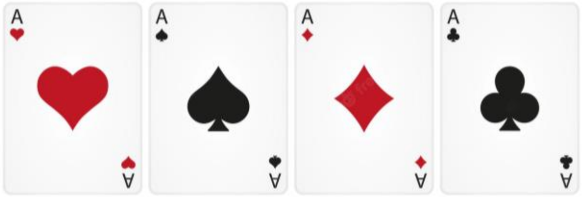
❖ Classical Probability

Example

Ad has drawn a card from a well-shuffled deck. Find the probability of some events

Drawing a king

Drawing a black card



Drawing a king

- A: "Drawing a king from a deck of cards"
- There are 52 cards in a deck of cards
  => $n(\Omega) = 52$
- There are 4 kings in a deck
  => $n(A) = 4$
  => $P(A) = 4/52 = 1/13$

Drawing a black card

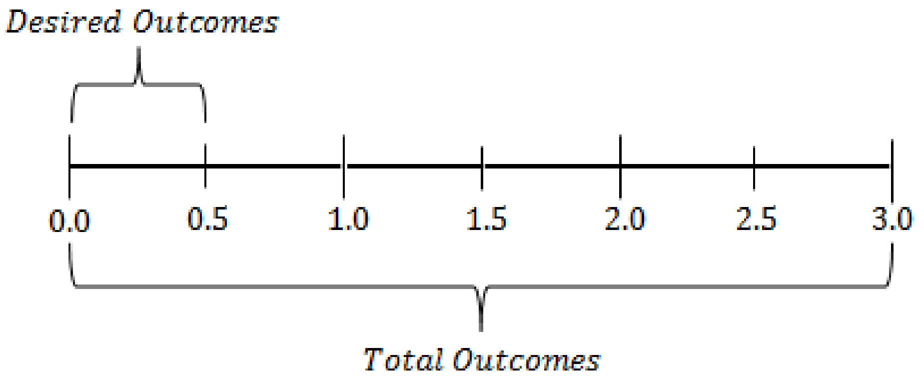
- A: "Drawing a black card from a deck of cards"
- There are 52 cards in a deck of cards
  => $n(\Omega) = 52$
- There are 26 black cards in a deck
  => $n(A) = 26$
  => $P(A) = 26/52 = 1/2$

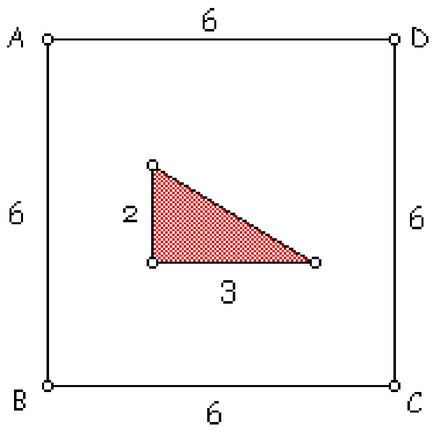
Probability

❖ Geometric Probability

When a variable is continuous, classical probability becomes impossible to "count" the outcomes.

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





X is a random real number between 0 and 3.

$$P(A) = \frac{length\ of\ segment\ where\ 0 < X < 0.5}{length\ of\ segment\ where\ 0 < X < 3} = \frac{0.5}{3} = \frac{1}{6}$$

Probability

❖ Geometric Probability

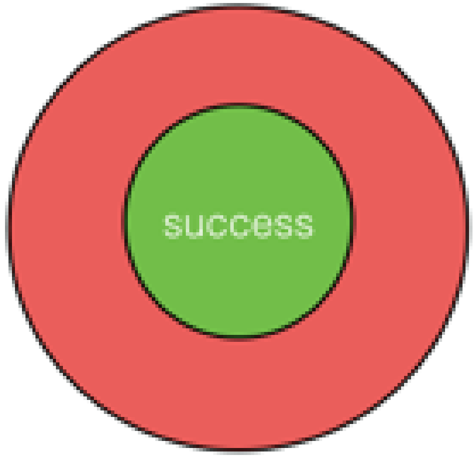
A dart is thrown at a circular dartboard such that it will land randomly over the area of the dartboard.

What is the probability that it lands closer to the center "success" than to the edge?

=> A: "A dart is thrown at the center area"

=> Measure: area in this 2D case:

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



Rules of probability

❖ The additive rule

Mutually exclusive events

$$P(A+B) = P(A) + P(B)$$

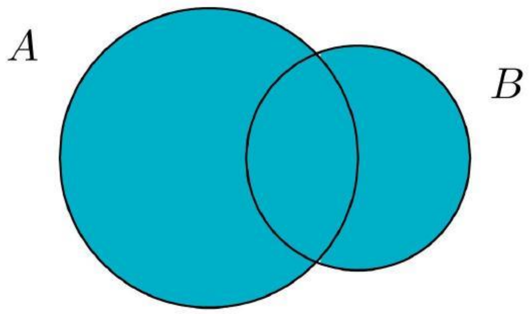
$$P(A \text{ or } B) = P(A) + P(B)$$

where A and B are mutually exclusive

In general

$$P(A+B) = P(A) + P(B) - P(AB)$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$



Example

Rolling a fair dice. What is the probability of A = {1, 5}?

- The die is fair => all six possible outcomes are equally likely

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$$

- The events {1},…,{6} are disjoint

$$1 = P(S) = P(\{1\}) + P(\{2\}) + \ldots + P(\{6\}) = 6P(\{1\})$$

$$\Rightarrow P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$$

- Since {1} and {5} are disjoint

$$\Rightarrow P(A) = P(\{1, 5\}) = P(\{1\}) + P(\{5\}) = 2/6 = 1/3$$

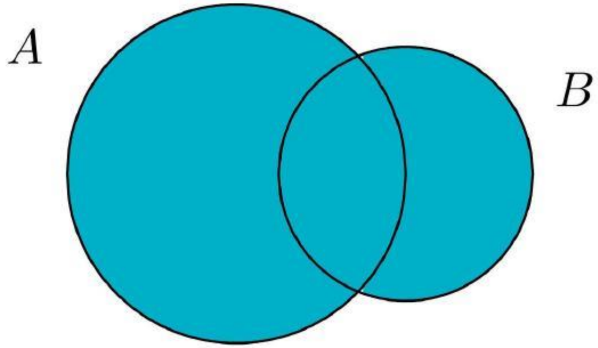
Rules of probability

❖ Example

Suppose we have the following information:

1. There is a 60 percent that Ad visits Ha Noi.
2. There is a 50 percent that Ad visits Ho Chi Minh.
3. There is a 30 percent that Ad visits 2 cities: Ha Noi and Ho Chi Minh

Find the probability that Ad visits Ha Noi or Ho Chi Minh?



Let's define

A: "Ad visits Ha Noi" => $P(A) = 0.6$

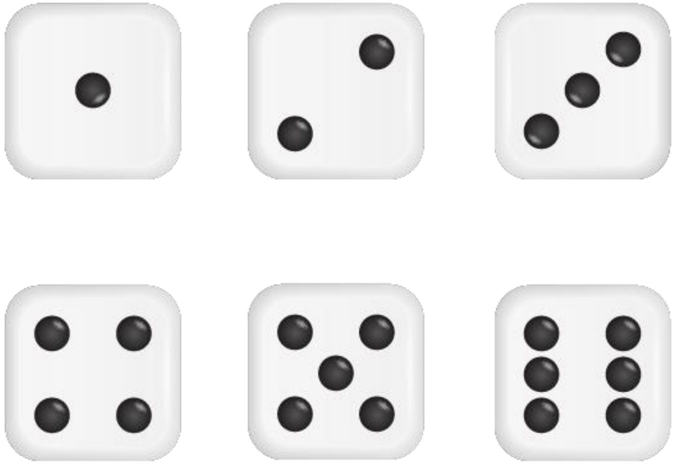
B: "Ad visits Ho Chi Minh" => $P(B) = 0.5$

$P(A \text{ and } B) = 0.3$

=> $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.6 + 0.5 - 0.3 = 0.8$

Rules of probability

❖ Complements



For any event A

$$P(A^c) = 1 - P(A)$$

$$P(A) = 1 - P(A^c)$$

Example

Find the probability that when we roll a dice we get a number different from 1 and 6?

Let's A: "Getting the number 1 and 6" => A = {1, 6}

"Getting a number different from 1 and 6" = A^c

Since $P(A) = P(1) + P(6) = 1/6 + 1/6 = 2/6 = 1/3$

P("Getting a number different than 1 and 6") = $1 - P(A) = 1 - 1/3 = 2/3$

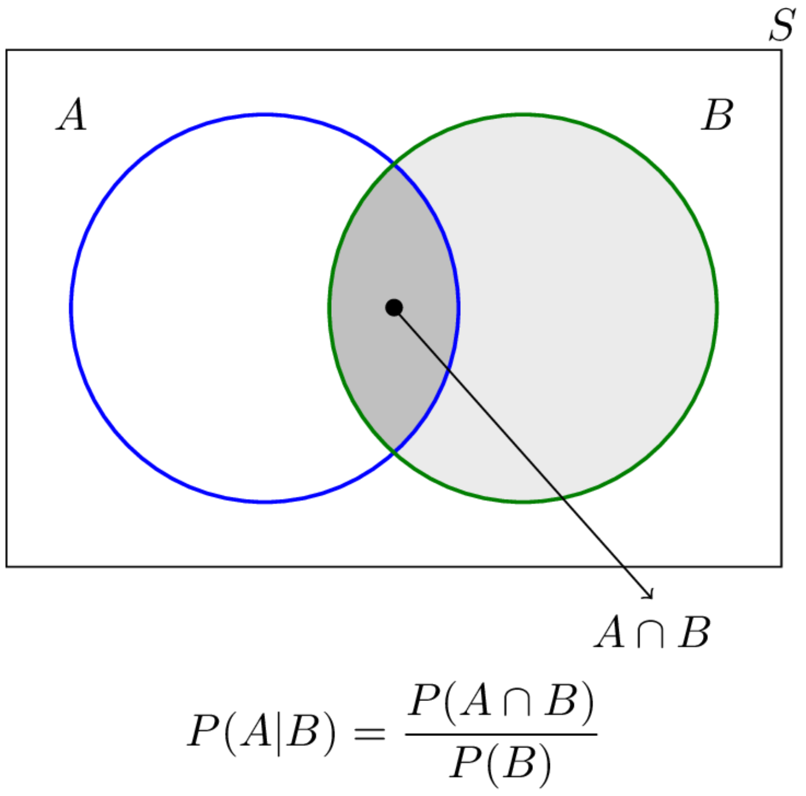
Conditional Probability

❖ Definition

Conditional Probability Formula

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

Probability that A occurs given that B has already occurred



A fair die is rolled

a) Find the probability that the number rolled is a five, given that it is odd.

b) Find the probability that the number rolled is odd, given that it is a five.

Conditional Probability

Example

Conditional probability formula

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

A fair dice is rolled

- Sample space S = {1, 2, 3, 4, 5, 6}, consisting of 6 equally likely outcomes
- A: "a five is rolled" => A = {5} => P(A) = 1/6
- B: "an odd number is rolled" => B = {1, 3, 5} => P(B) = 3/6 = 1/2

=> A and B = {5} => P(A and B) = 1/6

a) Find the probability that the number rolled is a five, given that it is odd.

P(A|B) = P(A and B)/P(B) = (1/6)/(1/2) = 1/3

b) Find the probability that the number rolled is odd, given that it is a five.

P(B|A) = P(B and A)/P(A) = P(A and B)/P(A) = (1/6)/(1/6) = 1

Multiplication Rule

Multiplication rule:

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

General:

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Example

In a factory there are 100 units of a certain product, 5 of which are defective.
We pick three units from the 100 units at random.

What is the probability that none of them are defective?

Multiplication Rule

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(An|A_1A_2\ldots A_{n-1})$$

Example

Let's A_i as the event i^{th} chosen unit is not defective, for i =1, 2, 3

=> Compute $P(A_1A_2A_3)$

$P(A_1) = 95/100$

Given that the first chosen item was good, the second item will be chosen from 94 good units and 5 defective units, thus: $P(A_2|A_1) = 94/99$

Given that the first and second chosen items were okay, the third item will be chosen from 93 good units and 5 defective units, thus: $P(A_3|A_2A_1) = 93/98$

=> $P(A_1A_2A3) = 95/100*94/99*93/98 = 0.8560$

Independent events

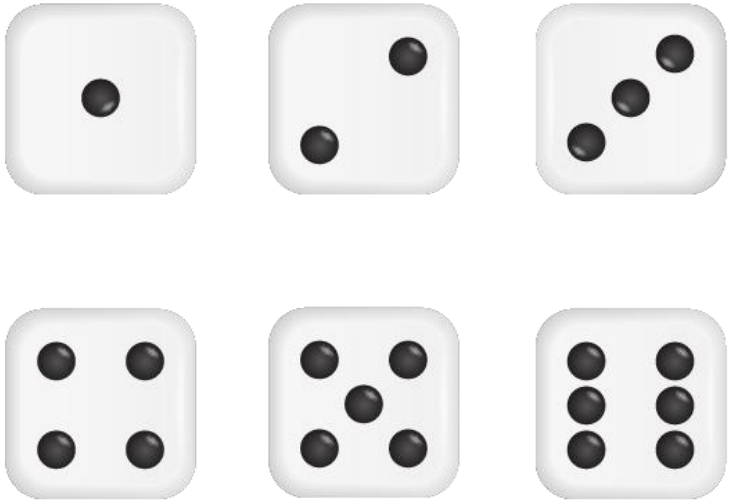
- Events A and B are independent if:

$$P(AB) = P(A)\,P(B)$$

- If A and B are not independent, they are dependent.

Roll a dice

Sample space: S = {1, 2, 3, 4, 5, 6}



Example

A single fair dice is rolled. Let A={3} and B={1,3,5}.

Are A and B independent?

Compute: $P(A) = 1/6$

$P(B) = 1/2$

$P(A \text{ and } B) = 1/6$

Since $P(A)P(B) = (1/6)*(1/2) = 1/12 \neq P(A \text{ and } B) = 1/6$

=> Events A and B: not independent

Bayes' Rule

For any two events A and B, where $P(A) \neq 0$:

LIKELIHOOD
The probability of "A" being True, given "B" True

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

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

POSTERIOR
The probability of "B" being True. Given "A" True

MARGINALIZATION
The probability of "A" being True.

Bayes' Rule

X is the result event, and $C_{1, 2,\ldots, n}$ are the causal events

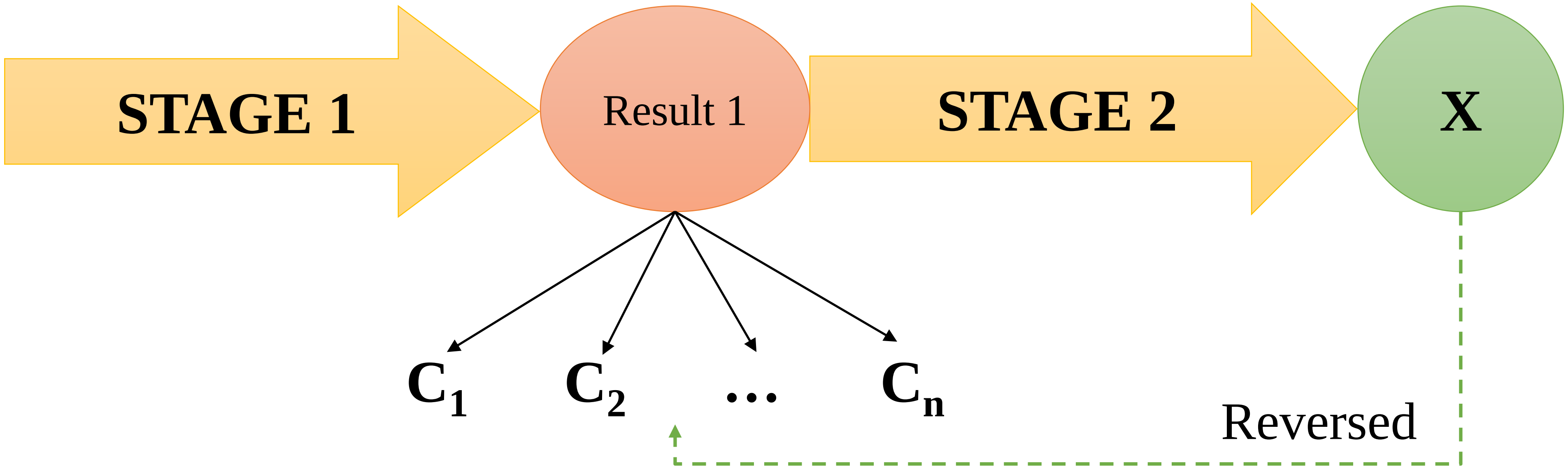
=> Know what causes occur

=> determine the probability that X will occur

Objective Bayes' rule:

Know the outcome X

Calculate the probability that the i^{th} cause occur: $P(C_i|X)$

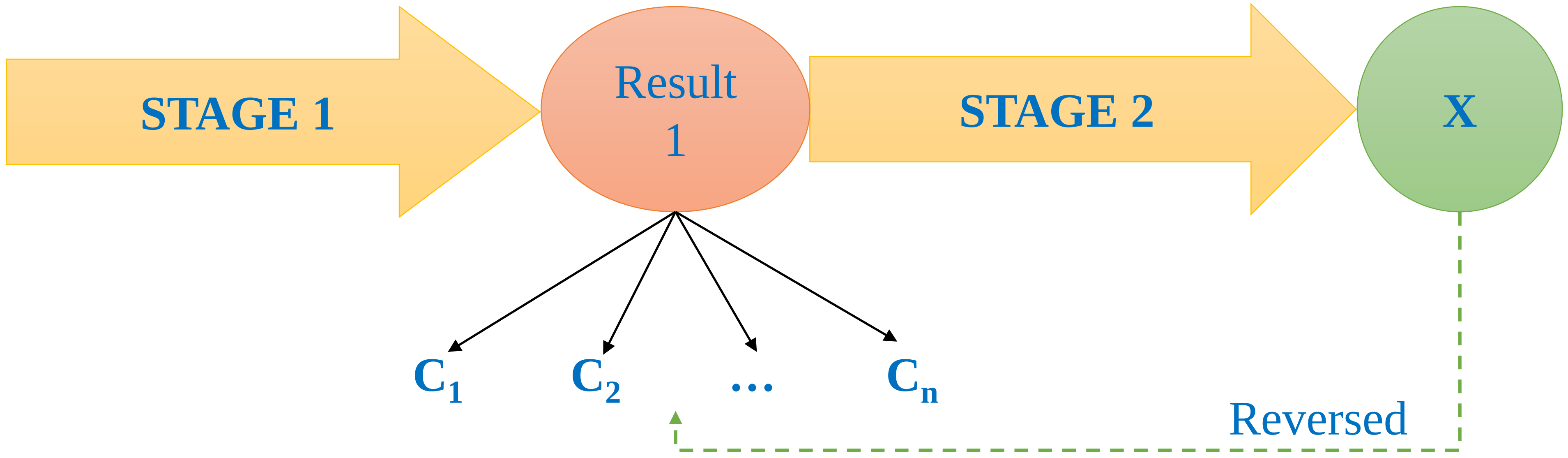


Bayes' Rule

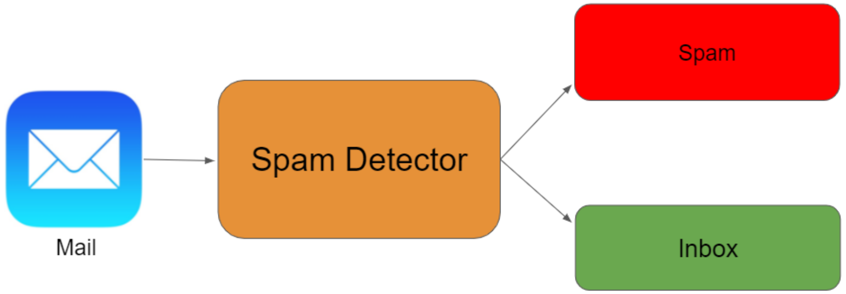
For any two events A and B, where $P(A) \neq 0$: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

If $C_1, C_2, .. C_n$: complete system of events and X is any event with $P(X) \neq 0$

$$P(C_i|X) = \frac{P(C_i)P(X|C_i)}{P(X)} = \frac{P(C_i)P(X|C_i)}{\sum_{j=1}^{n} P(C_j)P(X|C_j)}, i = 1, 2, \dots, n$$



Bayes' Rule



Example: Detect Spam E-Mail (Simple NLP problem)

Assume that the word 'offer' occurs in 80% of the spam messages in my account. Also, let's assume 'offer' occurs in 10% of my desired e-mails. If 30% of the received e-mails are considered as a spam. I will receive a new message which contains 'offer', what is the probability that it is spam?

Assume that I received 100 e-mails

100 e-mails

100*0.3=30 → Spam

100*0.7=70 → Desired

Spam:
- Contains 'offer' 30*0.8=24
- NOT contains 'offer' 30*0.2=6

Desired:
- Contains 'offer' 70*0.1=7
- NOT contains 'offer' 70*0.9=63

Bayes' Rule

Example: Detect Spam E-Mail (Simple NLP problem)

Let C_1: "Spam" and C_2: "Not spam"

=> C_1, C_2 : complete system of events

X: "contains the word 'offer'"

If a new message which contains 'offer', the probability that it is spam is:

$$P(C_1|X) = \frac{P(C_1)P(X|C_1)}{P(X)}$$

$P(C_1) = 0.3$; $P(C_2) = 1 - P(C_1) = 0.7$

$P(X|C_1) = 0.8$; $P(X|C_2) = 0.1$

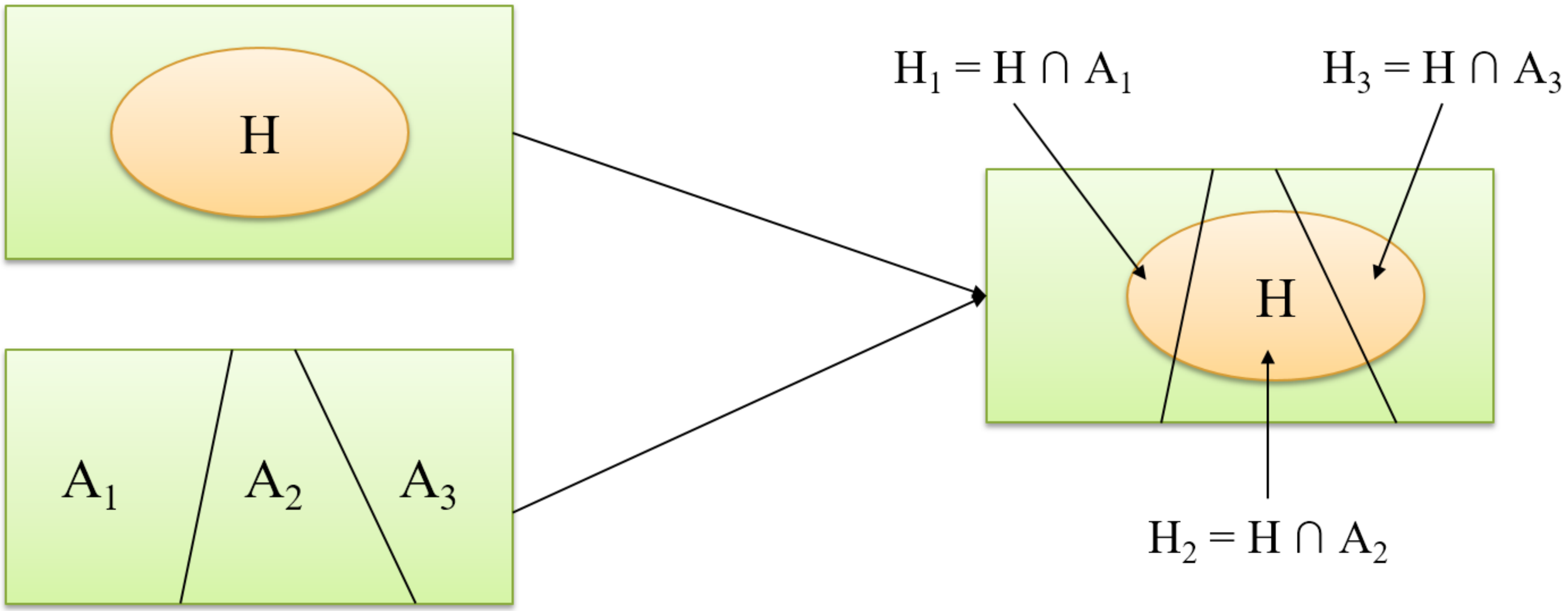
$P(X) = P(C_1)P(X|C_1) + P(C_2)P(X|C_2) = 0.3*0.8 + 0.7*0.1 = 0.31$

=> $P(C_1|X) = (0.8*0.3)/(0.31) = 0.774$

Total Probability Theorem

Let A_1, A_2, …, A_n – complete system of events and assume. Consider any event H such that H occurs only when one of the events $A_1, A_2, \ldots, A_n$ occurred

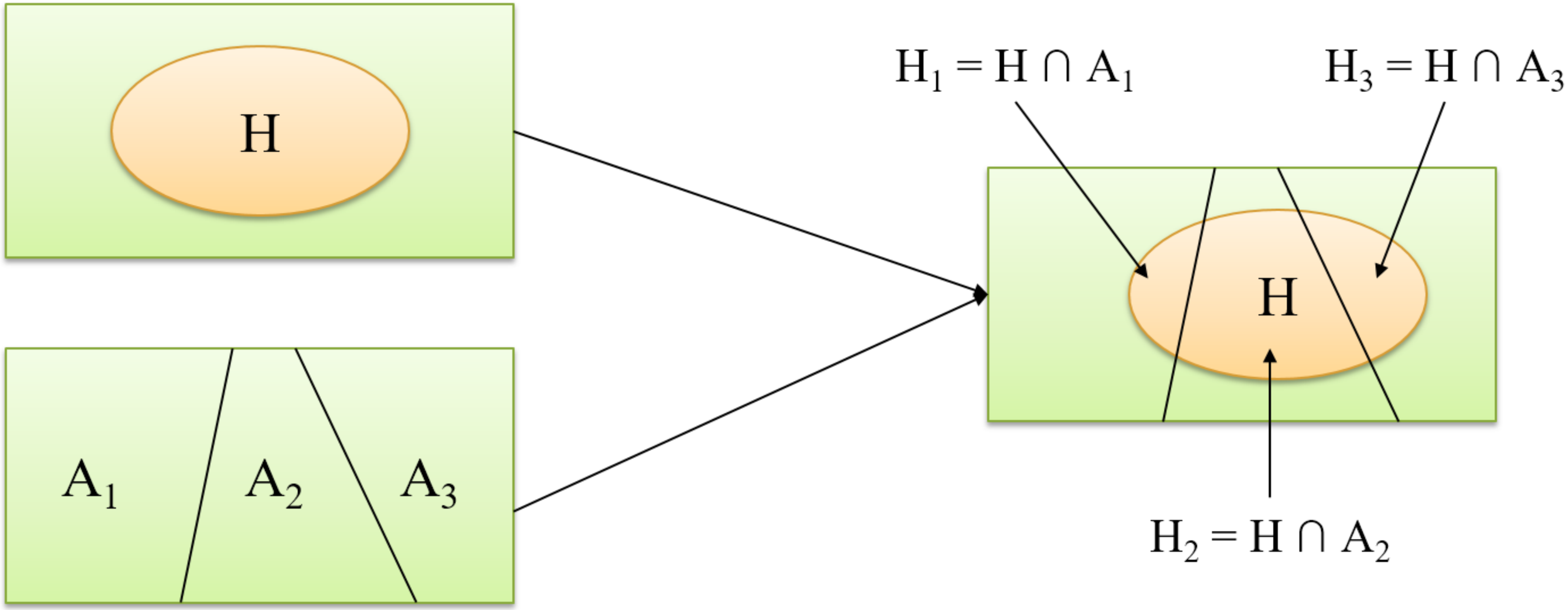
$$P(H) = P(H_1) + P(H_2) + P(H_3)$$
$$= P(A_1).P(H|A_1) + P(A_2).P(H|A_2) + P(A_3).P(H|A_3)$$



Total Probability Theorem

In general

$$P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$$



Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Example

Company M supplies 80% of widgets for a car shop and only 1% of their widgets turn out to be defective. Company N supplies the remaining 20% of widgets for the car shop and 3% of their widgets turn out to be defective. If a customer randomly purchases a widget from the car shop, what is the probability that it will be defective?

Company M	Supplies 80% of widgets 1% are defective
Company N	Supplies 20% of widgets 3% are defective

Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Company M	Supplies 80% of widgets 1% are defective
Company N	Supplies 20% of widgets 3% are defective

Example

H: "Widget being defective"

A_M: "Widget came from company M"

A_N: "Widget came from company N"

Events A_M and A_N : complete system of events

=> $P(A_M) = 0.8$; $P(A_N) = 0.2$; $P(H|A_M) = 0.01$; $P(H|A_N) = 0.03$

The probability that it will be defective:

$P(H) = P(H|A_M).\ P(A_M) + P(H|A_N).\ P(A_N)$

$= 0.01*0.8 + 0.03*0.2 = 0.014$

Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Example

I have three bags that each contain 100 marbles:

- Bag 1 has 75 red and 25 blue marbles
- Bag 2 has 60 red and 40 blue marbles
- Bag 3 has 45 red and 55 blue marbles.

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Example

H: “the chosen marble marble is red”

A_i : the event that I choose Bag I

=> $P(H|A_1) = 0.75$; $P(H|A_2) = 0.6$; $P(H|A_3) = 0.45$

Each bag contain 100 marbles and because their union is the entire sample space as one the bags will be chosen for sure, $P(A_1 \cup A_2 \cup A_3) = 1$

So, the probability that the chosen marble is red:

$$P(H) = P(A_1).P(H|A_1) + P(A_2).P(H|A_2) + P(A_3).P(H|A_3)$$

$$= 1/3*0.75 + 1/3*0.60 + 1/3*0.45 = 0.60$$

Outline

- Basic Concepts
- Bayes' Theorem
- Bernoulli Naïve Bayes Classifier
- Gaussian Naïve Bayes Classifier
- Histogram Equalization

Random Variable

❖ Bernoulli Random variables

A numerical description of the outcome of a statistical experiment

$$p(x) = p\{X = x\} = \begin{cases} p & when\ x = 1 \\ 1 - p & when\ x = 0 \end{cases}$$

Toss a coin

Sample space: S = {heads, tails}

Bernoulli Naïve Bayes Classifier

❖ Example: One feature

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

One feature: Studies

Two classes: Fail and Pass

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

Let C and X are random variables

$$p(c|x) = \frac{p(x|c) * p(c)}{p(x)}$$

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

Bernoulli Naïve Bayes Classifier

❖ **Example: One feature**

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

$$p(C = c1|X = x) = ? \qquad p(C = c2|X = x) = ?$$

$$p(C = c2|X = x) = ?$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$p(res = pass) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = fail) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$p(res = pass) = \frac{3}{6} = 0.5$$

$$p(res = fail) = \frac{3}{6} = 0.5$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$p(res = pass) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = fail) = \frac{3}{6} = \frac{1}{2}$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(stud = yes)$$

$$= p(stud = yes|res = pass) * p(res = pass) + p(stud = yes|res = fail) * p(res = fail)$$

$$= \frac{2}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{2} = \frac{1}{2}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

$$p(res = pass) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = fail) = \frac{3}{6} = \frac{1}{2}$$

$$p(stud = yes) = \frac{1}{2}$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$= \frac{2}{3} * \frac{1}{2} * \frac{2}{1} = \frac{2}{3}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

$$= \frac{1}{3} * \frac{1}{2} * \frac{2}{1} = \frac{1}{3}$$

Bernoulli Naïve Bayes Classifier

❖ Example: Three features

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

Three features: Confident, Studies, and Sick

Two classes: Fail and Pass

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

Let C and X are random variables

$$p(c|x) = \frac{p(x|c) * p(c)}{p(x)}$$

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

Bernoulli Naïve Bayes Classifier

❖ Example: Three features

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

$$p(C = c1|X = x) = ? \qquad p(C = c2|X = x) = ?$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

$$p(res = pass \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes|res = pass) * p(res = pass)}{p(conf = yes, stud = yes, sick = yes)}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes|res = fail) * p(res = fail)}{p(conf = yes, stud = yes, sick = yes)}$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

$$\boxed{p(res = pass) = \frac{3}{6} = 0.5}$$

$$\boxed{p(res = fail) = \frac{3}{6} = 0.5}$$

$$p(res = pass \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes|res = pass) * p(res = pass)}{p(\ conf = yes, stud = yes, sick = yes)}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes|res = fail) * p(res = fail)}{p(\ conf = yes, stud = yes, sick = yes)}$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail

$$p(res = pass) = \frac{3}{6} = 0.5$$

$$p(res = fail) = \frac{3}{6} = 0.5$$

$$p(conf = yes \mid res = pass) = \frac{2}{3}$$

$$p(stud = yes \mid res = pass) = \frac{2}{3}$$

$$p(sick = yes \mid res = pass) = \frac{1}{3}$$

$$p(conf = yes, stud = yes, sick = yes \mid res = pass)$$

$$= p(conf = yes \mid res = pass) * p(stud = yes \mid res = pass) * p(sick = yes \mid res = pass)$$

$$= \frac{2}{3} * \frac{2}{3} * \frac{1}{3} = \frac{4}{27}$$

$$p(res = pass \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes \mid res = pass) * p(res = pass)}{p(conf = yes, stud = yes, sick = yes)}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes) =$$

$$\frac{p(conf = yes, stud = yes, sick = yes \mid res = fail) * p(res = fail)}{p(conf = yes, stud = yes, sick = yes)}$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail

$$p(res = pass) = \frac{3}{6} = 0.5$$

$$p(res = fail) = \frac{3}{6} = 0.5$$

$$p(conf = yes \mid res = fail) = \frac{1}{3}$$

$$p(stud = yes \mid res = fail) = \frac{1}{3}$$

$$p(sick = yes \mid res = fail) = \frac{2}{3}$$

$$p(conf = yes, stud = yes, sick = yes \mid res = fail)$$

$$= p(conf = yes \mid res = fail) * p(stud = yes \mid res = fail) * p(sick = yes \mid res = fail)$$

$$= \frac{1}{3} * \frac{1}{3} * \frac{2}{3} = \frac{2}{27}$$

$$p(res = pass \mid conf = yes, stud = yes, sick = yes) = \frac{p(conf = yes, stud = yes, sick = yes \mid res = pass) * p(res = pass)}{p(conf = yes, stud = yes, sick = yes)}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes) = \frac{p(conf = yes, stud = yes, sick = yes \mid res = fail) * p(res = fail)}{p(conf = yes, stud = yes, sick = yes)}$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail

$$p(res = pass) = \frac{3}{6} = 0.5$$

$$p(res = fail) = \frac{3}{6} = 0.5$$

$$p(conf = yes) = \frac{3}{6} = \frac{1}{2}$$

$$p(stud = yes) = \frac{3}{6} = \frac{1}{2}$$

$$p(sick = yes) = \frac{3}{6} = \frac{1}{2}$$

$$p(conf = yes, stud = yes, sick = yes)$$

$$= p(conf = yes, stud = yes, sick = yes|res = pass) * p(res = pass)$$

$$+ p(conf = yes, stud = yes, sick = yes|res = fail) * p(res = fail)$$

$$= \frac{4}{27} * \frac{1}{2} + \frac{2}{27} * \frac{1}{2} = \frac{3}{27}$$

$$p(res = pass \mid conf = yes, stud = yes, sick = yes) = \frac{p(conf = yes, stud = yes, sick = yes|res = pass) * p(res = pass)}{p(conf = yes, stud = yes, sick = yes)}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes) = \frac{p(conf = yes, stud = yes, sick = yes|res = fail) * p(res = fail)}{p(conf = yes, stud = yes, sick = yes)}$$

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

$$p(res = pass) = \frac{3}{6} = 0.5 \qquad p(res = fail) = \frac{3}{6} = 0.5$$

$$p(conf = yes, stud = yes, sick = yes | res = pass) = \frac{4}{27}$$

$$p(conf = yes, stud = yes, sick = yes | res = fail) = \frac{2}{27}$$

$$p(conf = yes, stud = yes, sick = yes) = \frac{1}{12}$$

$$p(res = pass \mid conf = yes, stud = yes, sick = yes)$$

$$= \frac{p(conf = yes, stud = yes, sick = yes | res = pass) * p(res = pass)}{p(conf = yes, stud = yes, sick = yes)} = \frac{4}{27} * \frac{1}{2} * \frac{27}{3} = \frac{2}{3}$$

$$p(res = fail \mid conf = yes, stud = yes, sick = yes)$$

$$= \frac{p(conf = yes, stud = yes, sick = yes | res = fail) * p(res = fail)}{p(conf = yes, stud = yes, sick = yes)} = \frac{2}{27} * \frac{1}{2} * \frac{27}{3} = \frac{1}{3}$$

Outline

- Basic Concepts
- Bayes' Theorem
- Bernoulli Naïve Bayes Classifier
- Gaussian Naïve Bayes Classifier
- Histogram Equalization

Distribution

❖ Definition

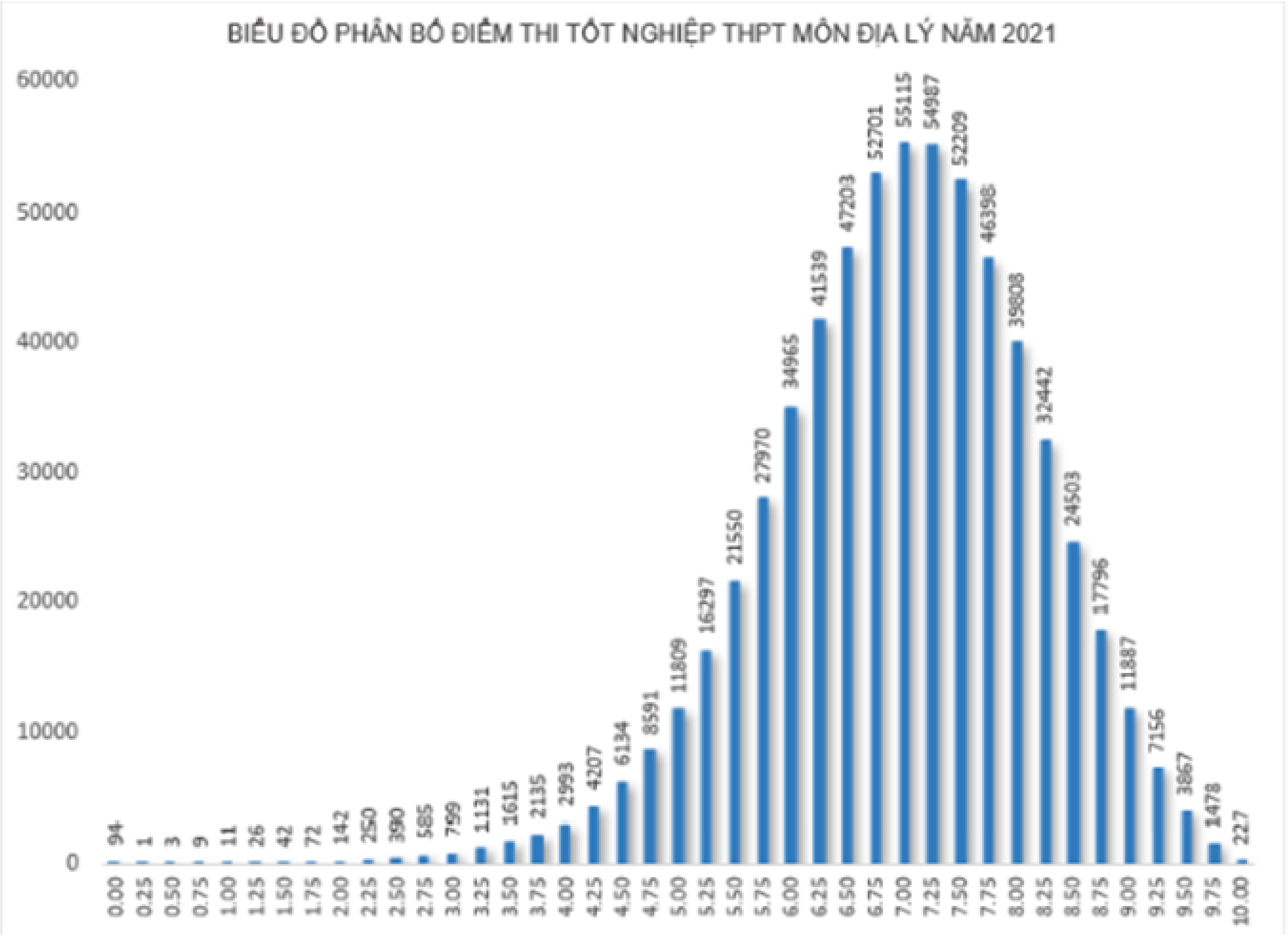
A distribution is simply a collection of data on a variable

Data are arranged in order

Statistics in Plain English

The probability distribution for a random variable describes how the probabilities are distributed over the values of the random variable.

https://www.britannica.com/science/statistics/Random-variables-and-probability-distributions



https://moet.gov.vn/tintuc/Pages/tin-tong-hop.aspx?ItemID=7451

Gaussian Distribution

Công thức

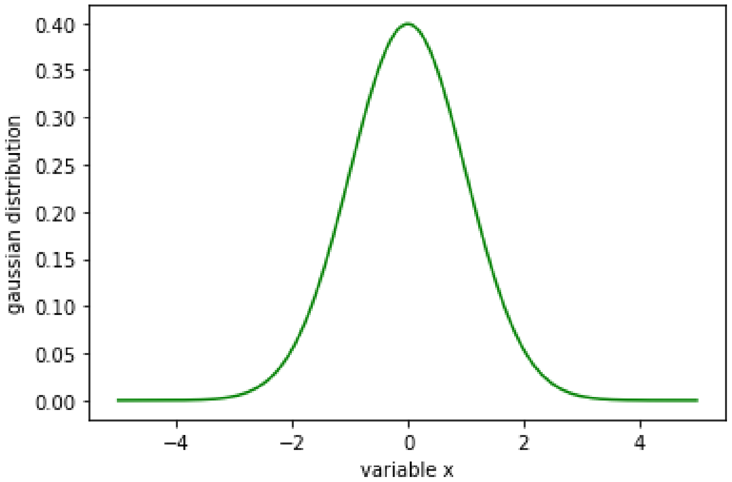
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$-\infty < x < \infty$

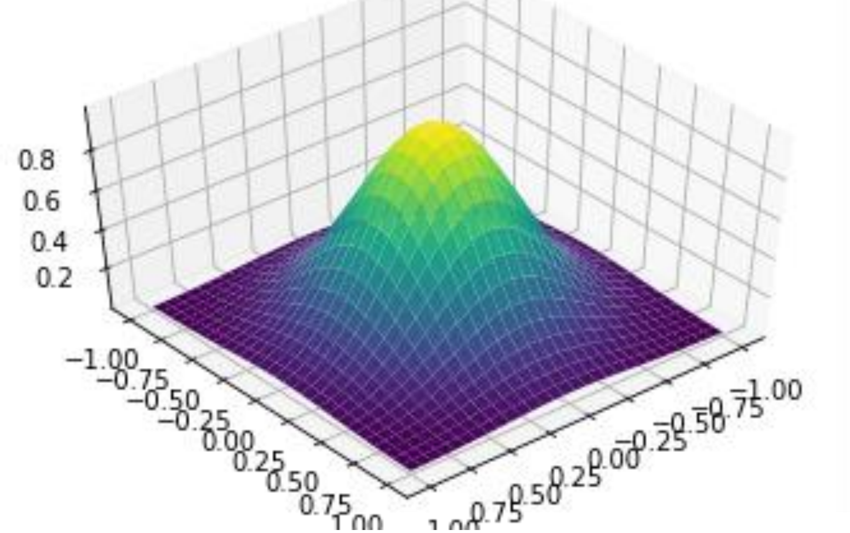
μ: mean

σ^2: variance

Đồ thị Gaussian 1D



Đồ thị Gaussian 2D



Ví dụ Cho $\mu = 0$ và $\sigma^2 = 1$. Tính giá trị hàm Gaussian với những giá trị x sau

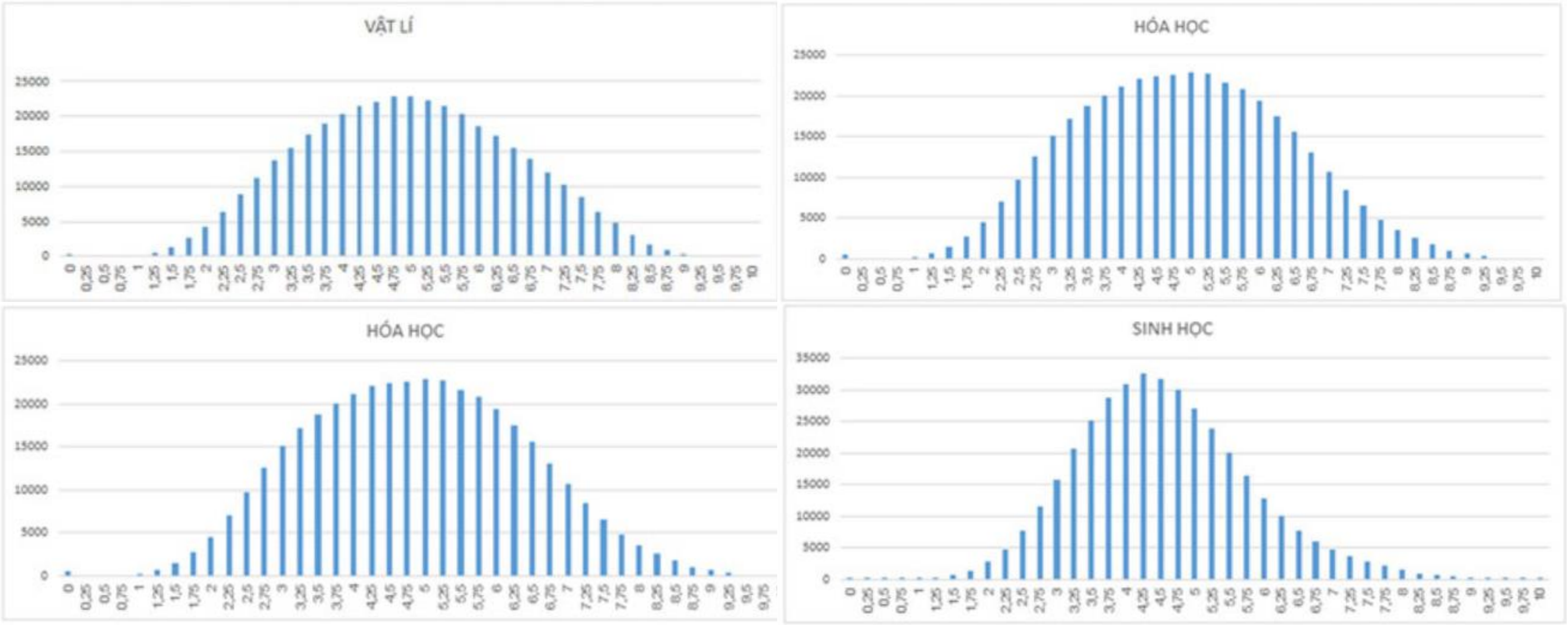
$$f(x = -4) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2*1}(-4-0)^2} = \frac{e^{-8}}{\sqrt{2\pi}} = \frac{1}{e^8\sqrt{2\pi}} = 1.3\text{e}{-04}$$

$x =$	[-4	-3	-2	-1	0	1	2	3]
$f(x) =$	[1.3e-04	4.4e-03	5.3e-02	2.4e-01	3.9e-01	2.4e-01	5.3e-02	4.4e-03]

Gaussian Distribution

Central limit theorem

Các giá trị sinh ra với số lượng đủ lớn từ một biến ngẫu nhiên độc lập xấp xỉ với phân bố Gaussian.

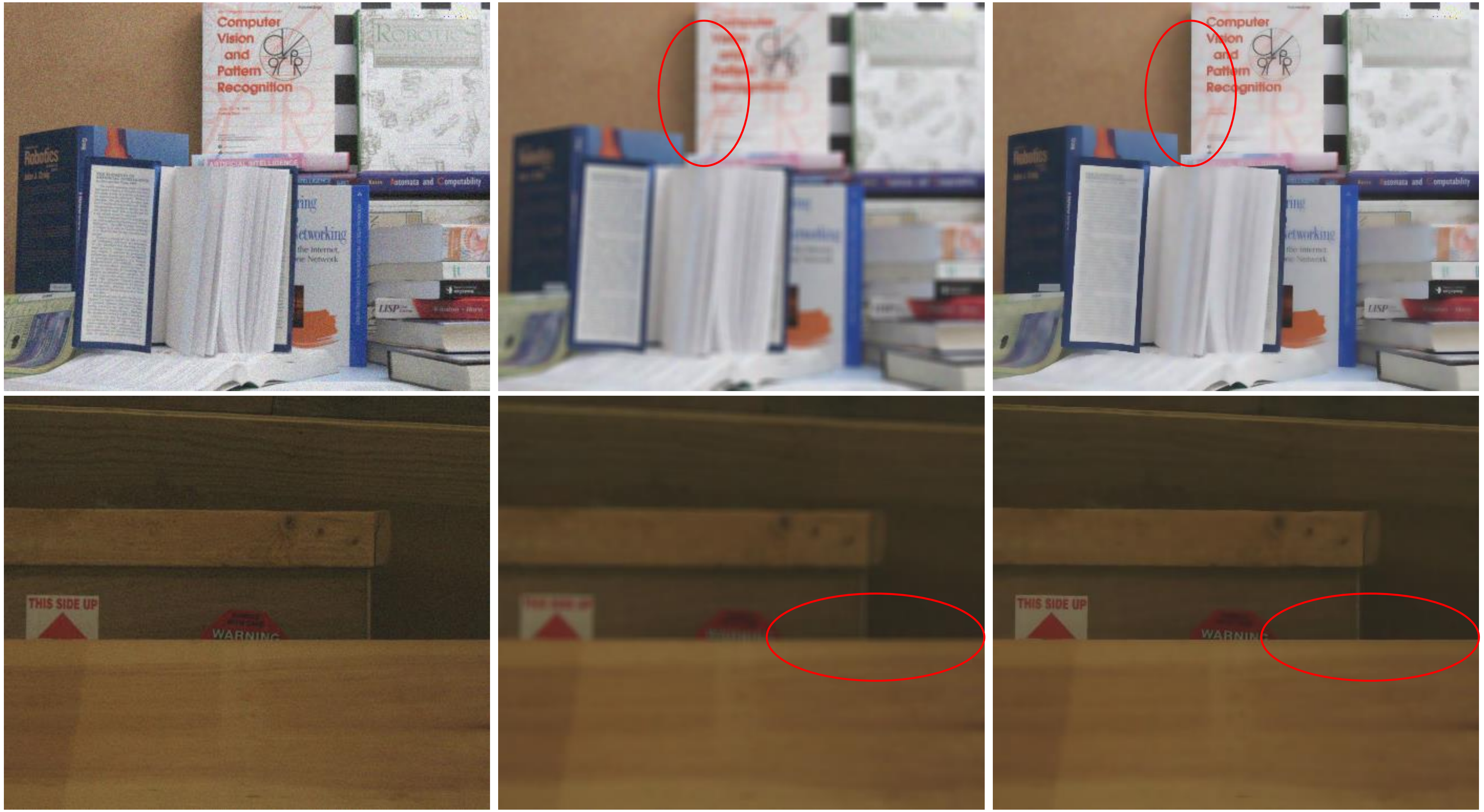


https://thanhnien.vn/giao-duc/bo-gd-dt-cong-bo-phan-tich-pho-diem-thi-thpt-quoc-gia-2018-981978.html

Bốn histogram trên là điểm thi của bốn môn trong đợt thi phổ thông quốc gia 2018. Chúng ta có thể quan sát thấy rằng phân bố của cả bốn điểm thi xấp xỉ phân bố Gaussian.

Ứng dụng Gaussian distribution cho image smoothing

Dùng hàm Gaussian để tính trọng số cho một kernel nxn.
Bilateral filter: dùng thêm thông tin pixel color để hoạt động tốt hơn cho chỗ viền/cạnh.



Ảnh gốc bị nhiễu | Smoothing với trọng số Gausian | Smoothing với trọng số Gausian+color

Gaussian Naïve Bayes Classifier

❖ Example: one feature

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

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

$$-\infty < x < \infty$$

μ: mean

σ^2: variance

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

Let C and X are random variables

$$p(c|x) = \frac{\boxed{p(x|c)} * p(c)}{p(x)}$$

One feature: Length (continuous)

Two classes: ‘0’ and ‘1’

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

class=? when Length=3.0

Gaussian Naïve Bayes Classifier

❖ Example

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

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

$-\infty < x < \infty$

μ: mean

σ^2: variance

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

One feature: Length (continuous)

Two classes: ‘0’ and ‘1’

class=? when Length=3.0

$$p(Cat = c|Len = 3.0) = \frac{\boxed{p(Len = 3.0|Cat = c)} * p(Cat = c)}{p(Len = 3.0)}$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

$\mu_0 = 1.52$

$\sigma_0^2 = 0.1416$

$$f(x|0) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}}$$

$\mu_1 = 3.88$

$\sigma_1^2 = 0.0776$

$$f(x|1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

Length	pdf(Length)	Category
1.4	1.007	0
1	0.408	0
1.3	0.893	0
1.9	0.636	0
2	0.469	0
3.8	1.374	1
4.1	1.048	1
3.9	1.428	1
4.2	0.74	1
3.4	0.324	1

$$p(Cat = 0|Len = 3.0) = \frac{p(Len = 3.0|Cat = 0) * p(Cat = 0)}{p(Len = 3.0)}$$

$$p(Cat = 1|Len = 3.0) = \frac{p(Len = 3.0|Cat = 1) * p(Cat = 1)}{p(Len = 3.0)}$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

$\mu_0 = 1.52$

$\sigma_0^2 = 0.1416$

$$f(x|0) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}}$$

$\mu_1 = 3.88$

$\sigma_1^2 = 0.0776$

$$f(x|1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$p(Cat = 0) = 0.5$

$p(Cat = 1) = 0.5$

$$p(Len = 3.0|Cat = 0) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(3.0-\mu_0)^2}{2\sigma_0^2}} = 0.0004638$$

$$p(Len = 3.0|Cat = 1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(3.0-\mu_1)^2}{2\sigma_1^2}} = 0.0097495$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

$$p(Cat = 0) = 0.5$$

$$p(Cat = 1) = 0.5$$

$$p(Len = 3.0|Cat = 0) = 0.0004638$$

$$p(Len = 3.0|Cat = 1) = 0.0097495$$

$$p(Len = 3.0) = p(Len = 3.0|Cat = 0) * p(Cat = 0) + p(Len = 3.0|Cat = 1) * p(Cat = 1)$$

$$= 0.0004638 * 0.5 + 0.0097495 * 0.5 = 0.005106$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Category
1.4	0
1	0
1.3	0
1.9	0
2	0
3.8	1
4.1	1
3.9	1
4.2	1
3.4	1

$$p(Cat = 0) = 0.5$$

$$p(Cat = 1) = 0.5$$

$$p(Len = 3.0|Cat = 0) = 0.0004638$$

$$p(Len = 3.0|Cat = 1) = 0.0097495$$

$$p(Len = 3.0) = 0.005106$$

$$p(Cat = 0|Len = 3.0) = \frac{p(Len = 3.0|Cat = 0) * p(Cat = 0)}{p(Len = 3.0)} = \frac{0.0004638 * 0.5}{0.005106} = 0.0455$$

$$p(Cat = 1|Len = 3.0) = \frac{p(Len = 3.0|Cat = 1) * p(Cat = 1)}{p(Len = 3.0)} = \frac{0.0097495 * 0.5}{0.005106} = 0.9545$$

Gaussian Naïve Bayes Classifier

❖ Example: two features

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

Two features: Length and Width (continuous)

Two classes: ‘0’ and ‘1’

class=? when Length=4.1 and Width=2.9

$$p(Cat = c|Len = 4.1, Wid = 2.9) = \frac{p(Len = 4.1, Wid = 2.9|Cat = c) * p(Cat = c)}{p(Len = 4.1, Wid = 2.9)}$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$\mu_{0l} = 4.867$ $\quad \sigma_{0l}^2 = 0.078$ $\quad \mu_{0w} = 3.367$ $\quad \sigma_{0w}^2 = 0.095$

$$f(Length|0) = \frac{1}{\sigma_{0l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0l})^2}{2\sigma_{0l}^2}} \qquad f(Width|0) = \frac{1}{\sigma_{0w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0w})^2}{2\sigma_{0w}^2}}$$

$\mu_{1l} = 6.216$ $\quad \sigma_{1l}^2 = 0.228$ $\quad \mu_{1w} = 2.916$ $\quad \sigma_{1w}^2 = 0.111$

$$f(Length|1) = \frac{1}{\sigma_{1l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1l})^2}{2\sigma_{1l}^2}} \qquad f(Width|1) = \frac{1}{\sigma_{1w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1w})^2}{2\sigma_{1w}^2}}$$

$$p(Cat = c|Len = 4.1, Wid = 2.9) = \frac{p(Len = 4.1, Wid = 2.9|Cat = c) * p(Cat = c)}{p(Len = 4.1, Wid = 2.9)}$$

Gaussian Naïve Bayes Classifier

❖ Example: two features

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$$f(Len = 4.1|0) = \frac{1}{\sigma_{0l}\sqrt{2\pi}} e^{-\frac{(4.1-\mu_{0l})^2}{2\sigma_{0l}^2}} = 0.0342$$

$$f(Wid = 2.9|0) = \frac{1}{\sigma_{0w}\sqrt{2\pi}} e^{-\frac{(2.9-\mu_{0w})^2}{2\sigma_{0w}^2}} = 0.4129$$

$$f(Len = 4.1|1) = \frac{1}{\sigma_{1l}\sqrt{2\pi}} e^{-\frac{(4.1-\mu_{1l})^2}{2\sigma_{1l}^2}} = 4e-5$$

$$f(Wid = 2.9|1) = \frac{1}{\sigma_{1w}\sqrt{2\pi}} e^{-\frac{(2.9-\mu_{1w})^2}{2\sigma_{1w}^2}} = 1.1938$$

$$p(Len = 4.1, Wid = 2.9|Cat = 0) = p(Len = 4.1\ |Cat = 0) * p(Wid = 2.9|Cat = 0)$$
$$= 0.0342 * 0.4129 = 0.01412$$

$$p(Len = 4.1, Wid = 2.9|Cat = 1) = p(Len = 4.1\ |Cat = 1) * p(Wid = 2.9|Cat = 1)$$
$$= 0.00004 * 1.1938 = 0.000047$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$$p(Len = 4.1, Wid = 2.9|Cat = 0) = 0.01412$$

$$p(Len = 4.1, Wid = 2.9|Cat = 1) = 0.000047$$

$$p(Cat = 0) = 0.5$$

$$p(Cat = 1) = 0.5$$

$$p(Len = 4.1, Wid = 2.9) = p(Len = 4.1, Wid = 2.9|Cat = 0)p(Cat = 0)$$

$$+ p(Len = 4.1, Wid = 2.9|Cat = 1)p(Cat = 1)$$

$$= 0.01412 * 0.5 + 0.000047 * 0.5 = 0.00708$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$$p(Len = 4.1, Wid = 2.9|Cat = 0) = 0.01412$$

$$p(Len = 4.1, Wid = 2.9|Cat = 1) = 0.000047$$

$$p(Cat = 0) = 0.5$$

$$p(Cat = 1) = 0.5$$

$$p(Len = 4.1, Wid = 2.9) = 0.00708$$

$$p(Cat = 0|Len = 4.1, Wid = 2.9) = \frac{p(Len = 4.1, Wid = 2.9|Cat = 0) * p(Cat = 0)}{p(Len = 4.1, Wid = 2.9)}$$

$$= \frac{0.01412 * 0.5}{0.00708} = 0.9962$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1

$$p(Len = 4.1, Wid = 2.9 | Cat = 0) = 0.01412$$

$$p(Len = 4.1, Wid = 2.9 | Cat = 1) = 0.000047$$

$$p(Cat = 0) = 0.5$$

$$p(Cat = 1) = 0.5$$

$$p(Len = 4.1, Wid = 2.9) = 0.00708$$

$$p(Cat = 0 | Len = 4.1, Wid = 2.9) = 0.9962$$

$$p(Cat = 1 | Len = 4.1, Wid = 2.9) = \frac{p(Len = 4.1, Wid = 2.9 | Cat = 1) * p(Cat = 1)}{p(Len = 4.1, Wid = 2.9)}$$

$$= \frac{0.000047 * 0.5}{0.00708} = 0.0038$$

Gaussian Naïve Bayes Classifier

❖ Example: Three classes

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

Two features: Length and Width (continuous)

Two classes: ‘0’, ‘1’, and ‘2’

class=? when Length=5.2 and Width=2.4

$$p(Cat = c|Len = 5.2, Wid = 2.4)$$

$$= \frac{p(Len = 5.2, Wid = 2.4|Cat = c) * p(Cat = c)}{p(Len = 5.2, Wid = 2.4)}$$

$$p(Cat = 0|Len = 5.2, Wid = 2.4)$$

$$p(Cat = 1|Len = 5.2, Wid = 2.4)$$

$$p(Cat = 2|Len = 5.2, Wid = 2.4)$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$\mu_{0l} = 4.867 \quad \sigma_{0l}^2 = 0.078 \quad \mu_{0w} = 3.367 \quad \sigma_{0w}^2 = 0.095$

$$f(Length|0) = \frac{1}{\sigma_{0l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0l})^2}{2\sigma_{0l}^2}} \qquad f(Width|0) = \frac{1}{\sigma_{0w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0w})^2}{2\sigma_{0w}^2}}$$

$\mu_{1l} = 6.216 \quad \sigma_{1l}^2 = 0.228 \quad \mu_{1w} = 2.916 \quad \sigma_{1w}^2 = 0.111$

$$f(Length|1) = \frac{1}{\sigma_{1l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1l})^2}{2\sigma_{1l}^2}} \qquad f(Width|1) = \frac{1}{\sigma_{1w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1w})^2}{2\sigma_{1w}^2}}$$

$\mu_{2l} = 6.216 \quad \sigma_{2l}^2 = 0.228 \quad \mu_{2w} = 2.916 \quad \sigma_{2w}^2 = 0.111$

$$f(Length|2) = \frac{1}{\sigma_{2l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{2l})^2}{2\sigma_{2l}^2}} \qquad f(Width|2) = \frac{1}{\sigma_{2w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{2w})^2}{2\sigma_{2w}^2}}$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = \frac{1}{\sigma_{0l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0l})^2}{2\sigma_{0l}^2}} = 0.7023$$

$$f(Wid = 2.4|0) = \frac{1}{\sigma_{0w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{0w})^2}{2\sigma_{0w}^2}} = 0.0097$$

$$f(Len = 5.2|1) = \frac{1}{\sigma_{1l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1l})^2}{2\sigma_{1l}^2}} = 0.0866$$

$$f(Wid = 2.4|1) = \frac{1}{\sigma_{1w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{1w})^2}{2\sigma_{1w}^2}} = 0.3606$$

$$f(Len = 5.2|2) = \frac{1}{\sigma_{2l}\sqrt{2\pi}} e^{-\frac{(x-\mu_{2l})^2}{2\sigma_{2l}^2}} = 0.6785$$

$$f(Wid = 2.4|2) = \frac{1}{\sigma_{2w}\sqrt{2\pi}} e^{-\frac{(x-\mu_{2w})^2}{2\sigma_{2w}^2}} = 0.1839$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$f(Len = 5.2|0) = 0.7023$ $\quad$ $f(Wid = 2.4|0) = 0.0097$

$f(Len = 5.2|1) = 0.0866$ $\quad$ $f(Wid = 2.4|1) = 0.3606$

$f(Len = 5.2|2) = 0.6785$ $\quad$ $f(Wid = 2.4|2) = 0.1839$

$$p(Len = 5.2, Wid = 2.4|Cat = 0)$$

$$= p(Len = 5.2\ |Cat = 0) * p(Wid = 2.4|Cat = 0)$$

$$= 0.7023 * 0.0097 = 0.00681$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = 0.7023 \qquad f(Wid = 2.4|0) = 0.0097$$

$$f(Len = 5.2|1) = 0.0866 \qquad f(Wid = 2.4|1) = 0.3606$$

$$f(Len = 5.2|2) = 0.6785 \qquad f(Wid = 2.4|2) = 0.1839$$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1)$$

$$= p(Len = 5.2\ |Cat = 1) * p(Wid = 2.4|Cat = 1)$$

$$= 0.0866 * 0.3606 = 0.03122$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = 0.7023 \qquad f(Wid = 2.4|0) = 0.0097$$

$$f(Len = 5.2|1) = 0.0866 \qquad f(Wid = 2.4|1) = 0.3606$$

$$f(Len = 5.2|2) = 0.6785 \qquad f(Wid = 2.4|2) = 0.1839$$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1) = 0.03122$$

$$p(Len = 5.2, Wid = 2.4|Cat = 2)$$

$$= p(Len = 5.2\,|Cat = 2) * p(Wid = 2.4|Cat = 2)$$

$$= 0.6785 * 0.1839 = 0.1247$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$f(Len = 5.2|0) = 0.7023$ $\qquad f(Wid = 2.4|0) = 0.0097$

$f(Len = 5.2|1) = 0.0866$ $\qquad f(Wid = 2.4|1) = 0.3606$

$f(Len = 5.2|2) = 0.6785$ $\qquad f(Wid = 2.4|2) = 0.1839$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1) = 0.03122$$

$$p(Len = 5.2, Wid = 2.4|Cat = 2) = 0.1247$$

$$p(Cat = 0) = 0.333$$

$$p(Cat = 1) = 0.333$$

$$p(Cat = 2) = 0.334$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = 0.7023 \qquad f(Wid = 2.4|0) = 0.0097$$

$$f(Len = 5.2|1) = 0.0866 \qquad f(Wid = 2.4|1) = 0.3606$$

$$f(Len = 5.2|2) = 0.6785 \qquad f(Wid = 2.4|2) = 0.1839$$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681 \qquad p(Cat = 0) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1) = 0.03122 \qquad p(Cat = 1) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 2) = 0.1247 \qquad p(Cat = 2) = 0.334$$

$$\begin{aligned} p(Len = 5.2, Wid = 2.4) = \ & p(Len = 5.2, Wid = 2.4|Cat = 0) * p(Cat = 0) \\ & + p(Len = 5.2, Wid = 2.4|Cat = 1) * p(Cat = 1) \\ & + p(Len = 5.2, Wid = 2.4|Cat = 2) * p(Cat = 2) \\ = \ & 0.00681 * 0.333 + 0.03122 * 0.333 + 0.1247 * 0.334 = 0.0543 \end{aligned}$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = 0.7023 \qquad f(Wid = 2.4|0) = 0.0097$$

$$f(Len = 5.2|1) = 0.0866 \qquad f(Wid = 2.4|1) = 0.3606$$

$$f(Len = 5.2|2) = 0.6785 \qquad f(Wid = 2.4|2) = 0.1839$$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681 \qquad p(Cat = 0) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1) = 0.03122 \qquad p(Cat = 1) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 2) = 0.1247 \qquad p(Cat = 2) = 0.334$$

$$p(Len = 5.2, Wid = 2.4) = 0.0543$$

$$p(Cat = 0|Len = 5.2, Wid = 2.4)$$

$$= \frac{p(Len = 5.2, Wid = 2.4|Cat = 0) * p(Cat = 0)}{p(Len = 5.2, Wid = 2.4)} = \frac{0.00681 * 0.333}{0.0543} = 0.041$$

Gaussian Naïve Bayes Classifier

❖ Example

Length	Width	Category
4.9	3	0
4.7	3.2	0
4.6	3.1	0
5	3.6	0
5.4	3.9	0
4.6	3.4	0
6.4	3.2	1
6.9	3.1	1
5.5	2.3	1
6.5	2.8	1
5.7	2.8	1
6.3	3.3	1
5.6	2.2	2
5.1	1.5	2
5.6	1.4	2
5.9	2.1	2
5.3	1.8	2
5.7	1.9	2

$$f(Len = 5.2|0) = 0.7023$$

$$f(Wid = 2.4|0) = 0.0097$$

$$f(Len = 5.2|1) = 0.0866$$

$$f(Wid = 2.4|1) = 0.3606$$

$$f(Len = 5.2|2) = 0.6785$$

$$f(Wid = 2.4|2) = 0.1839$$

$$p(Len = 5.2, Wid = 2.4|Cat = 0) = 0.00681$$

$$p(Cat = 0) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 1) = 0.03122$$

$$p(Cat = 1) = 0.333$$

$$p(Len = 5.2, Wid = 2.4|Cat = 2) = 0.1247$$

$$p(Cat = 2) = 0.334$$

$$p(Len = 5.2, Wid = 2.4) = 0.0543$$

$$p(Cat = 0|Len = 5.2, Wid = 2.4) = 0.042$$

$$p(Cat = 1|Len = 5.2, Wid = 2.4) = 0.192$$

$$p(Cat = 2|Len = 5.2, Wid = 2.4) = 0.766$$

Outline

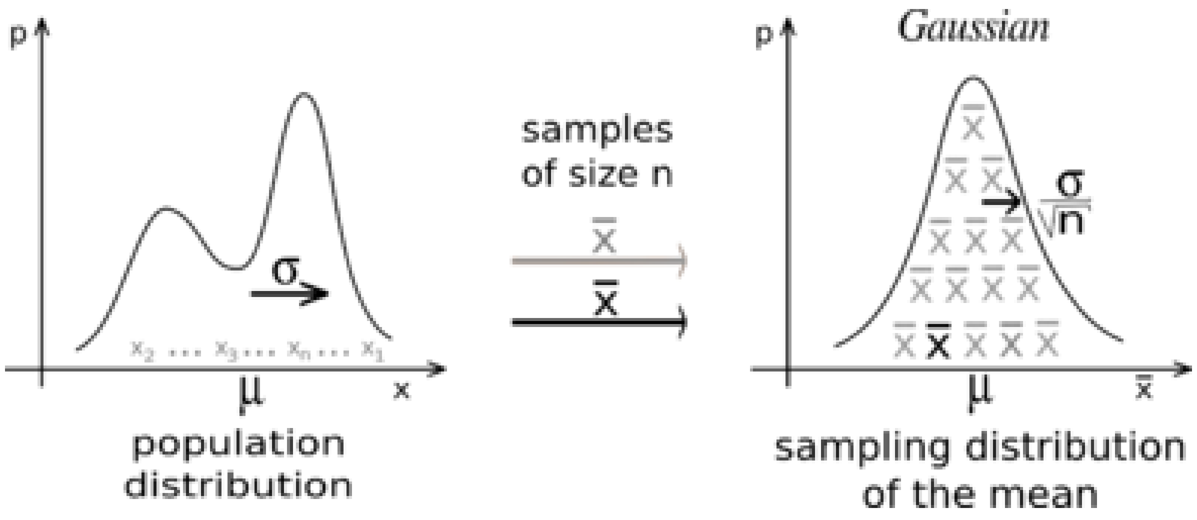
- Basic Concepts
- Bayes' Theorem
- Bernoulli Naïve Bayes Classifier
- Gaussian Naïve Bayes Classifier
- Histogram Equalization

Central limit theorem

❖ Example

When independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

https://en.wikipedia.org/wiki/Central_limit_theorem



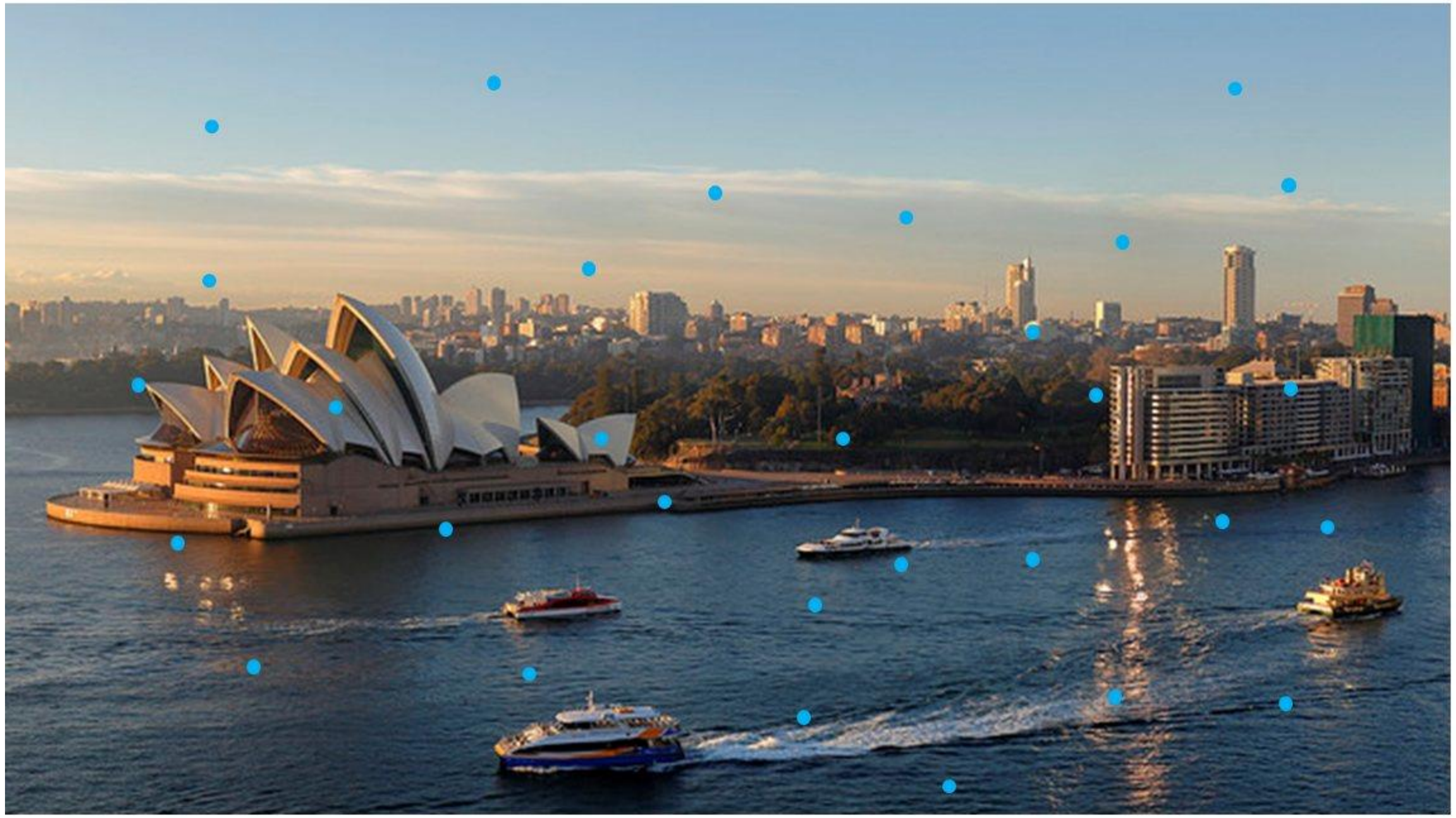
Central Limit Theorem

Randomly selected 30 pixels

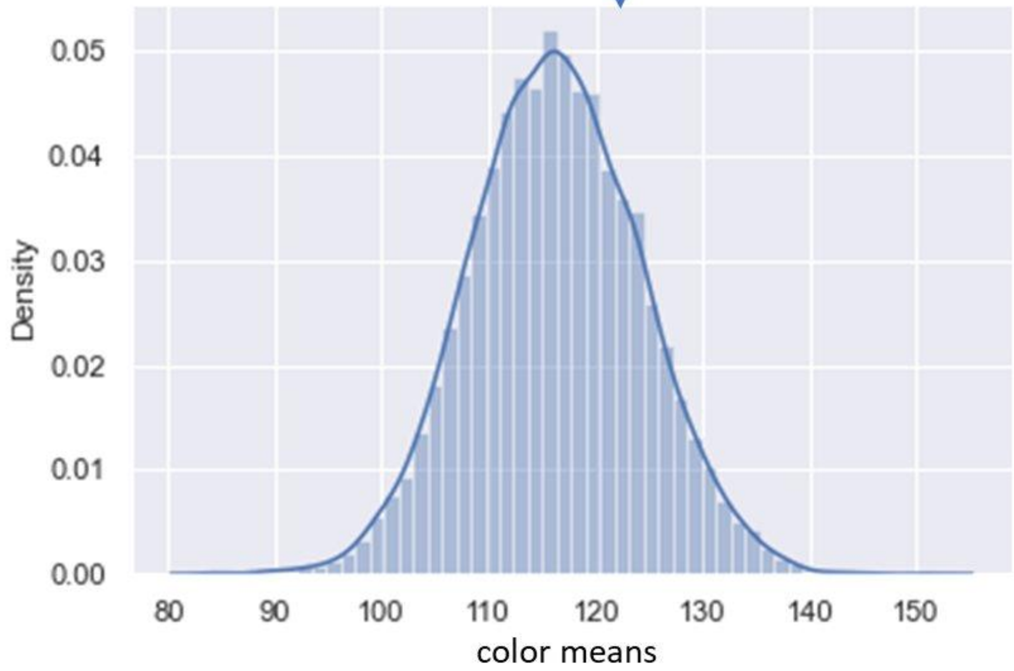
Compute color mean for the 30 pixels

Repeat 10000 times

Histogram of the 10000 color means



Population: pixel colors of the entire image



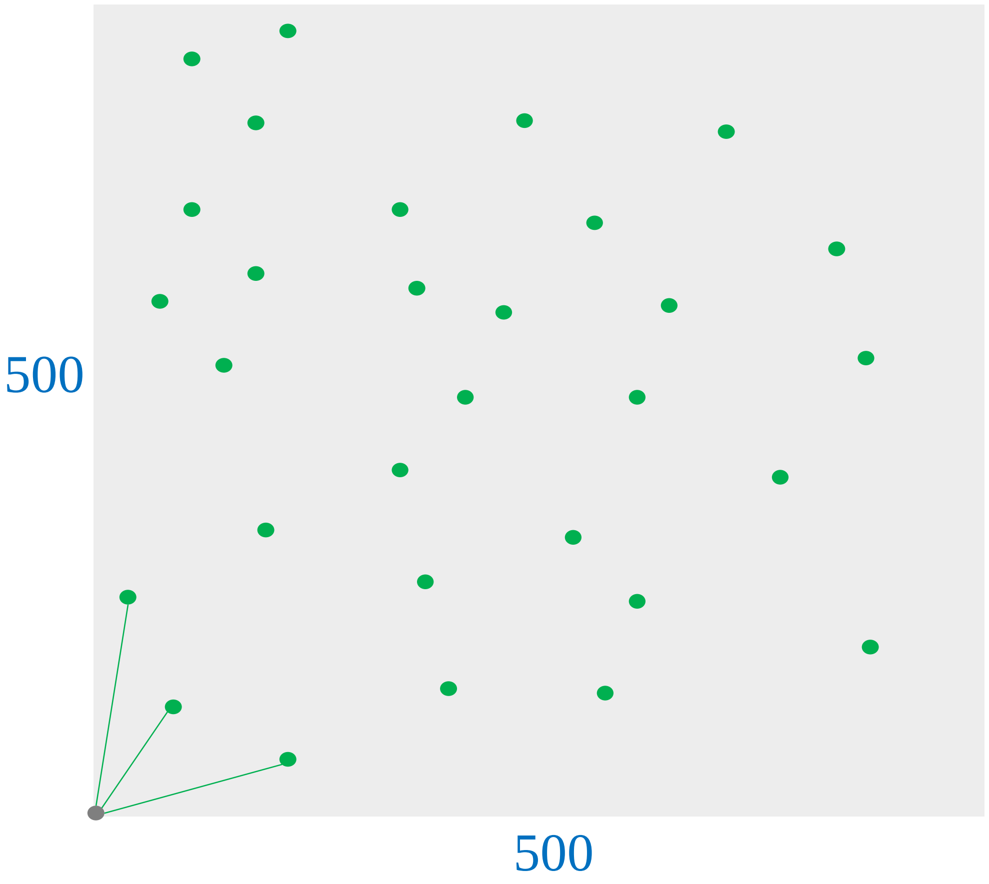
Example

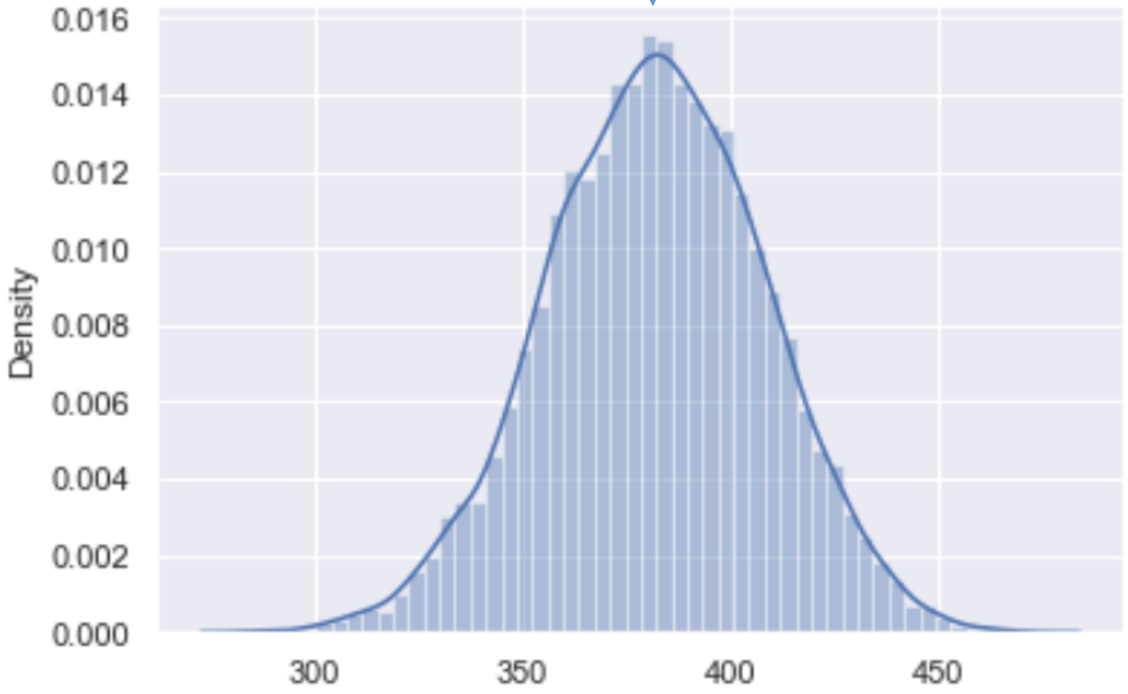
Randomly selected 30 pixels

Compute distance mean

Repeat 10000 times

Histogram of 1000 distance means





Classifier

❖ Example

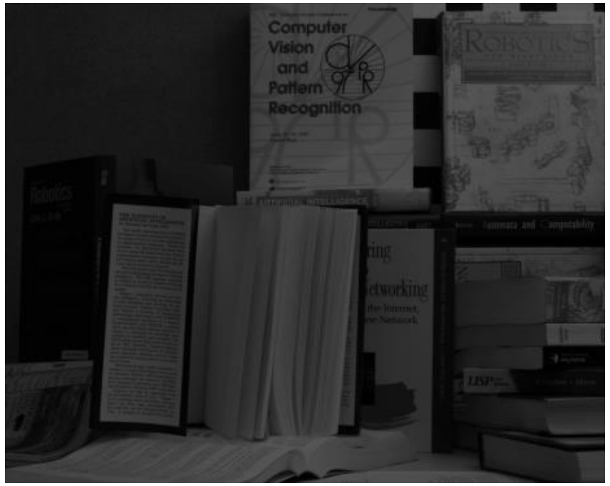
Outline

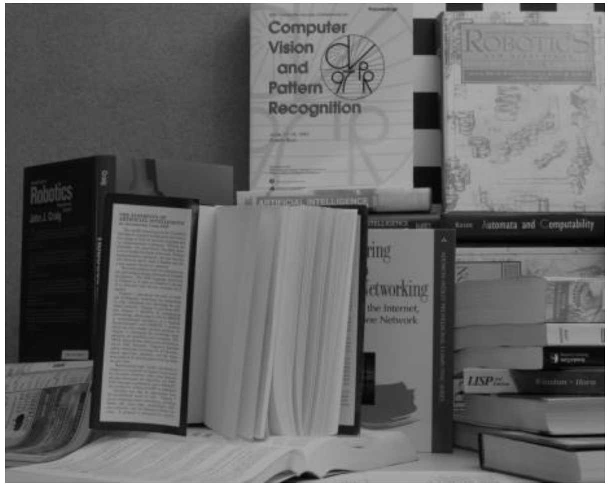
- Basic Concepts
- Bayes’ Theorem
- Bernoulli Naïve Bayes Classifier
- Gaussian Naïve Bayes Classifier
- Histogram Equalization

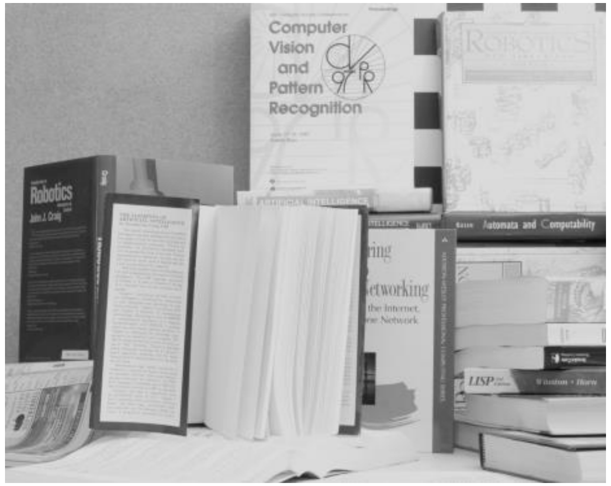
Histogram cho ảnh grayscale

Pixel có 256 giá trị khác nhau (từ 0 đến 255) → chia thành 256 bin. Mỗi bin chứa (đếm) số lượng pixel có cùng giá trị.

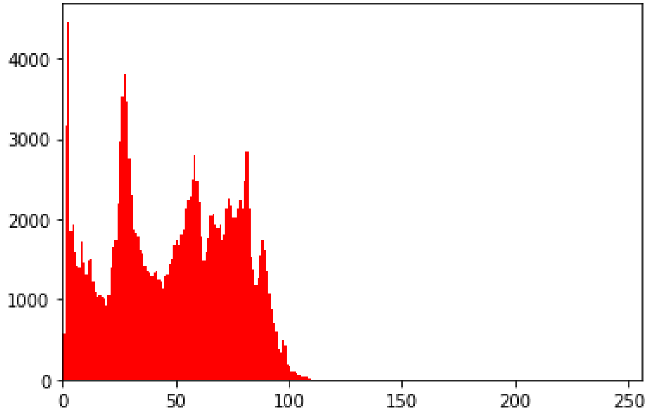
Ảnh



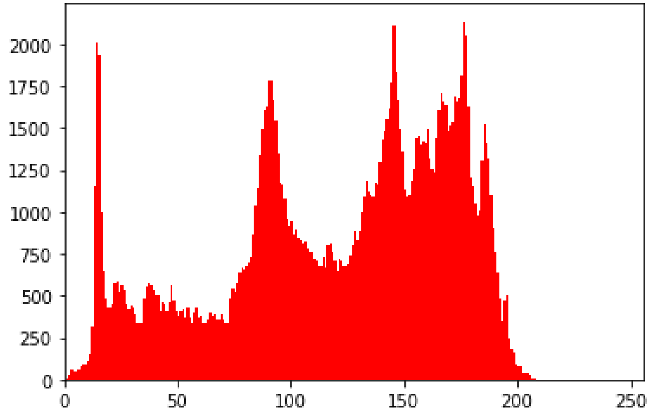




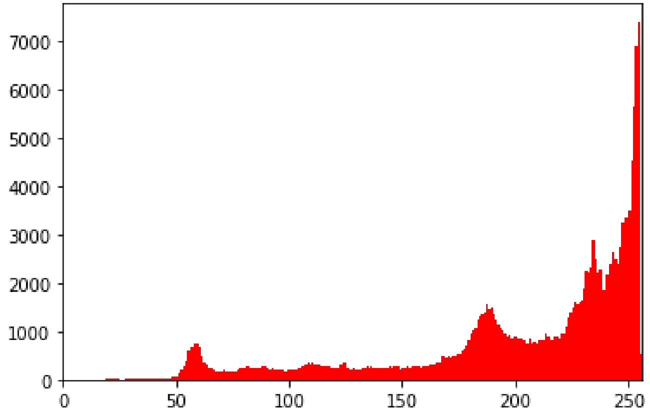
Histogram



256 bin

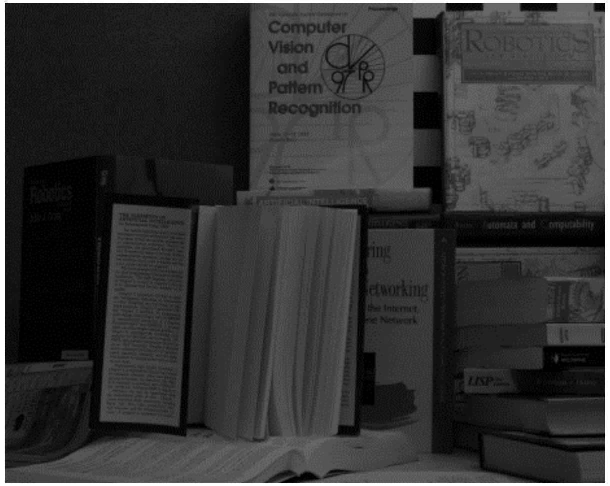


256 bin

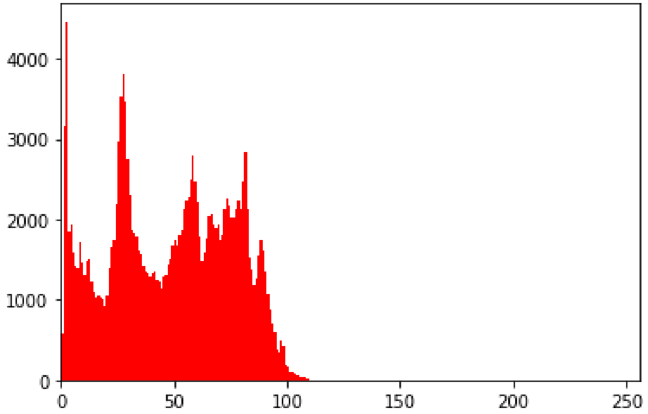


256 bin

Histogram cho ảnh grayscale



Ảnh

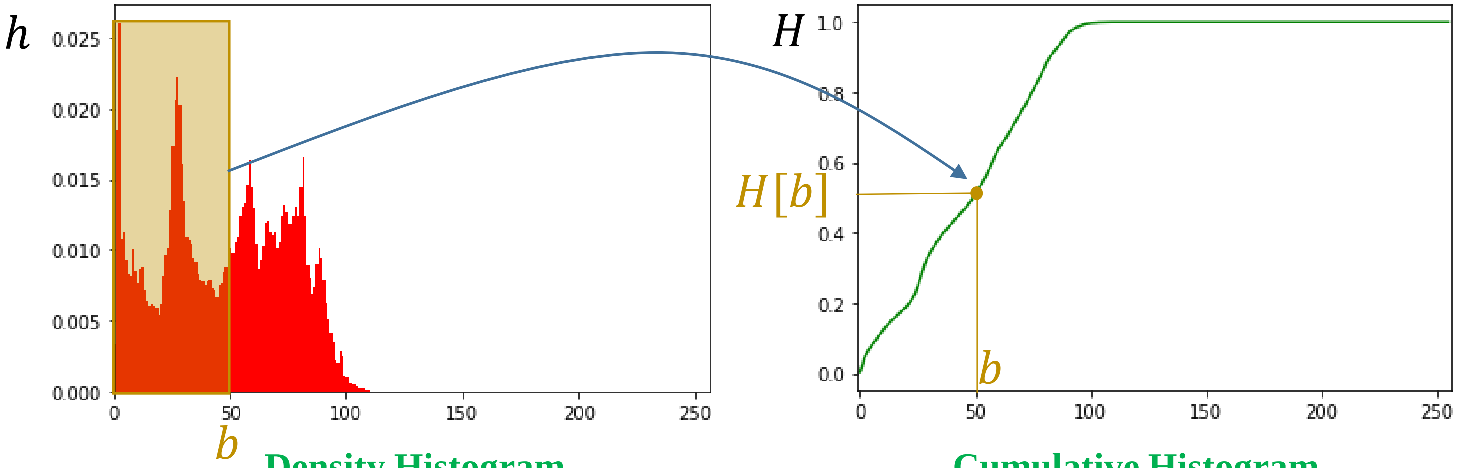


Histogram

Gọi N là tổng số pixel của ảnh và n_b là số lượng pixel ở bin thứ b.

Giá trị density histogram ở bin thứ b được tính như sau:

$$h[b] = \frac{n_b}{N}$$



Density Histogram

Cumulative Histogram

Giá trị cumulative histogram ở bin thứ b được tính như sau:

$$H[b] = \sum_{k=0}^{b} h[k]$$

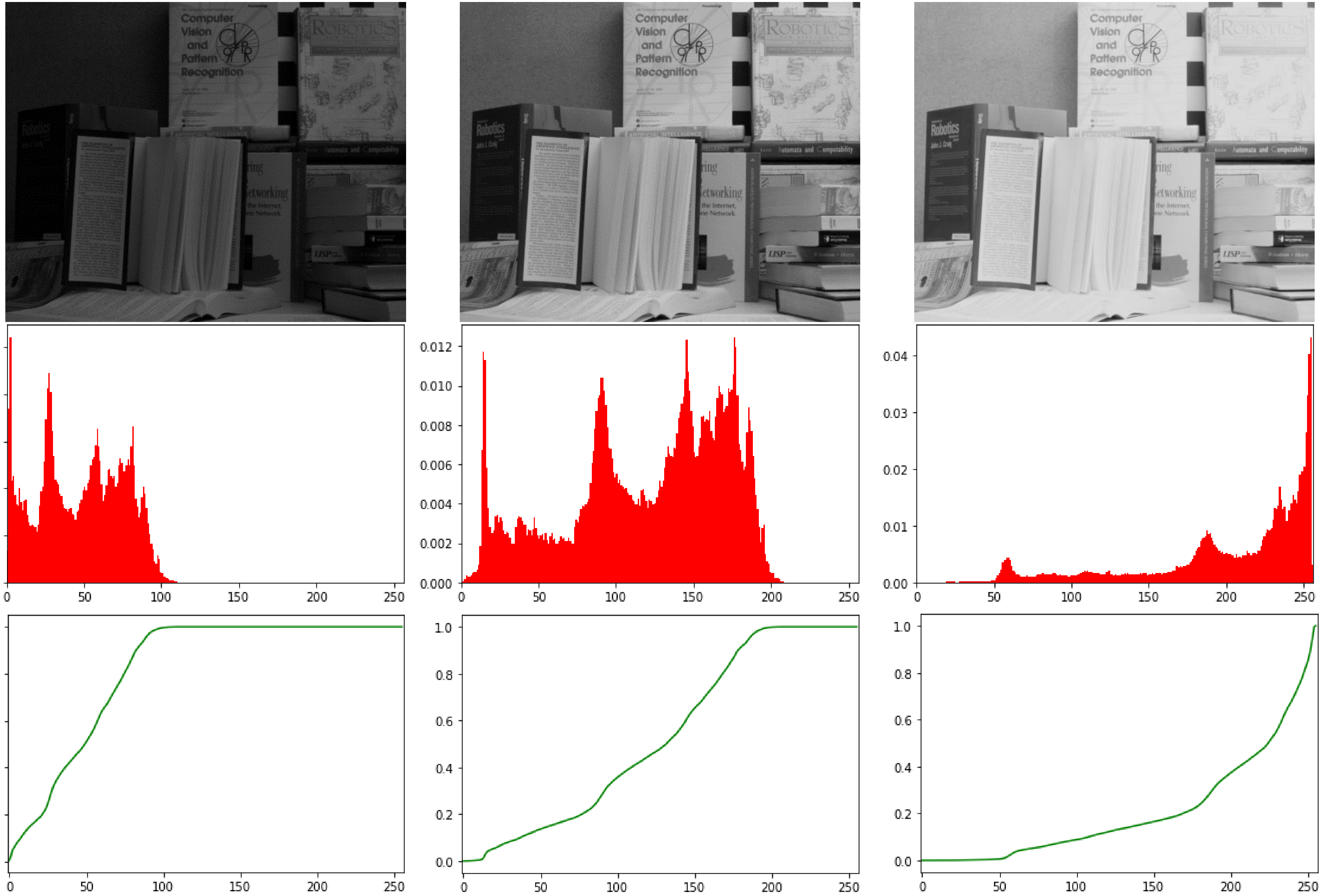
Cộng dồn giá trị $h[k]$ từ 0 đến b

Histogram cho ảnh grayscale

Ảnh

Density Histogram

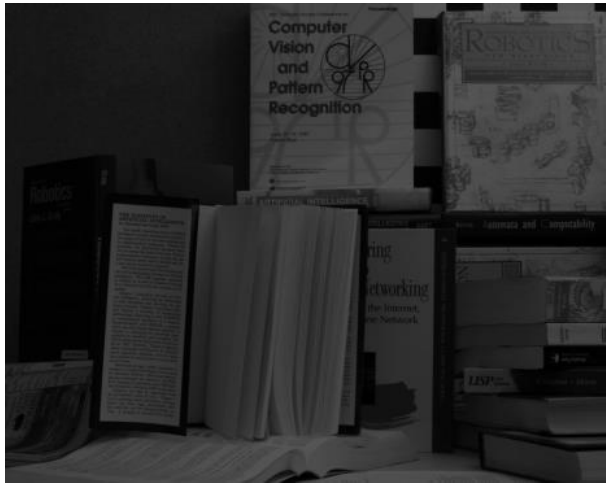
Cumulative Histogram



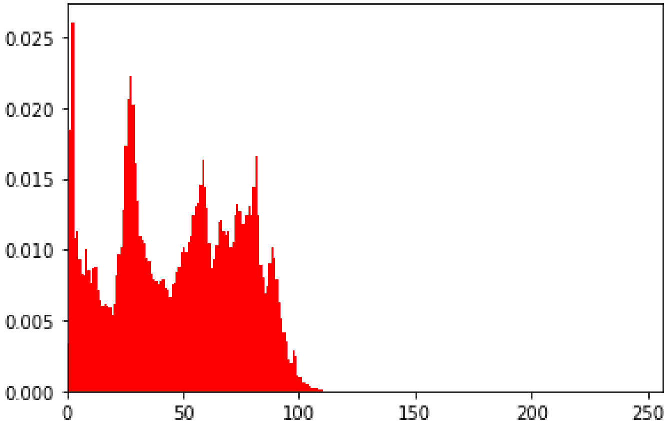
Histogram equalization

Được dùng để tăng độ tương phản của ảnh

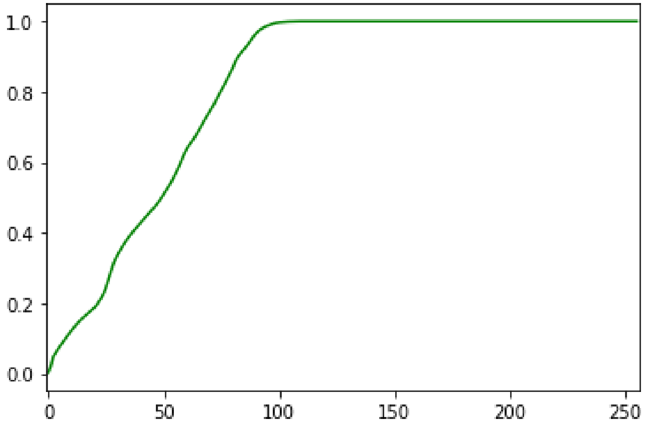
Idea: Kéo phân bố của density histogram sao cho xấp xỉ với uniform distribution.



Ảnh



Density Histogram

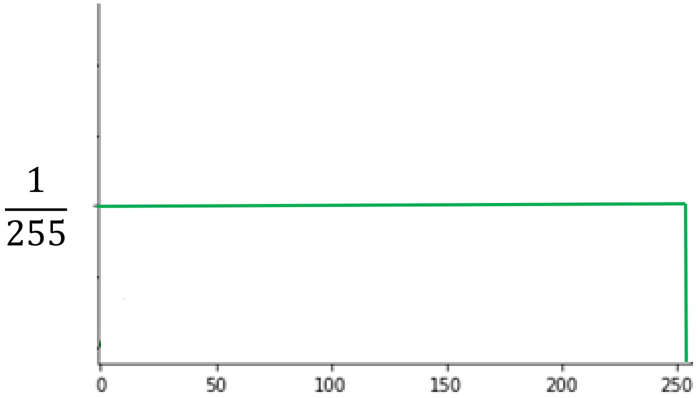


Cumulative Histogram

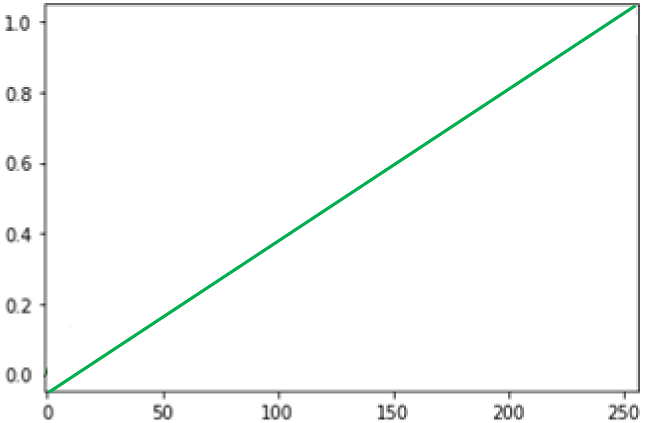
Công thức

$$b_{new} = \lfloor 255 * H[b] + 0.5 \rfloor$$

Các bin (giá trị pixel) được thay đổi theo $H[b]$

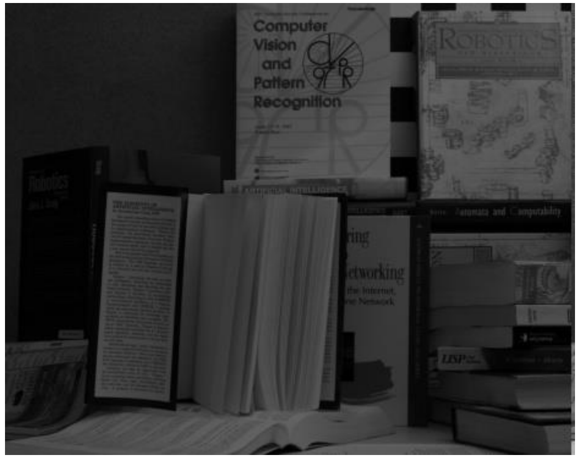


Uniform density function trong khoảng [0, 255]

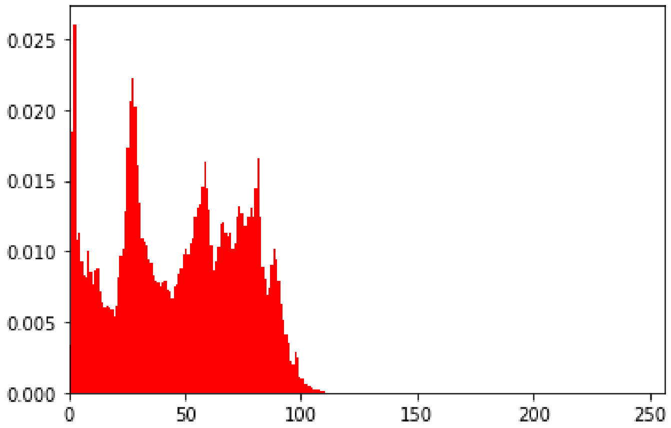


Cumulative density function cho uniform distribution

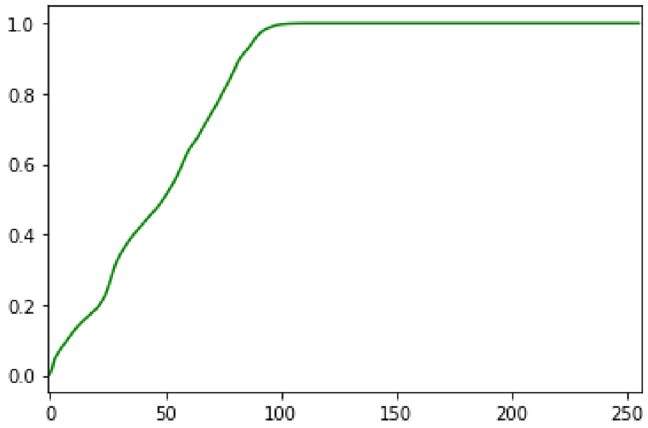
Histogram equalization để tăng độ tương phản



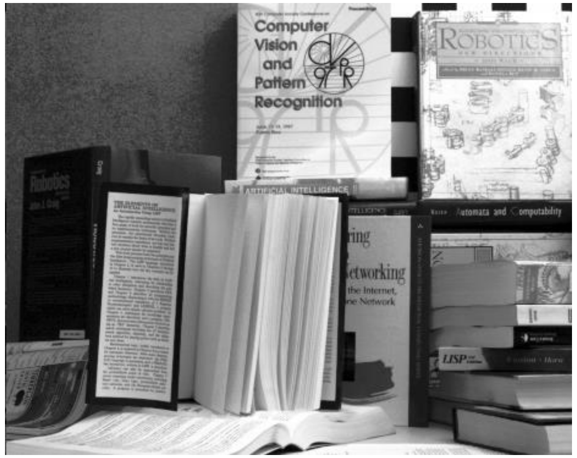
Ảnh gốc



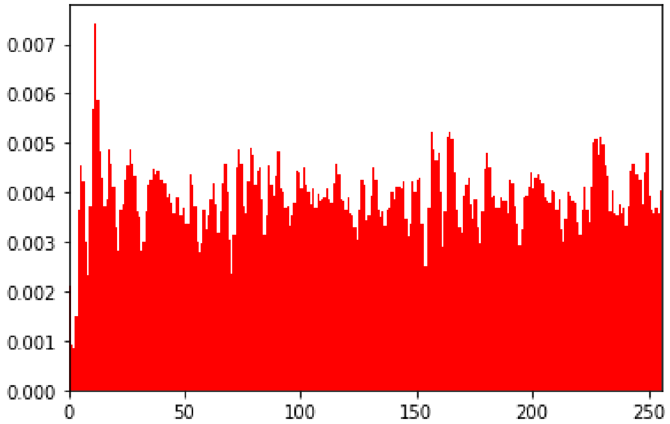
Density Histogram



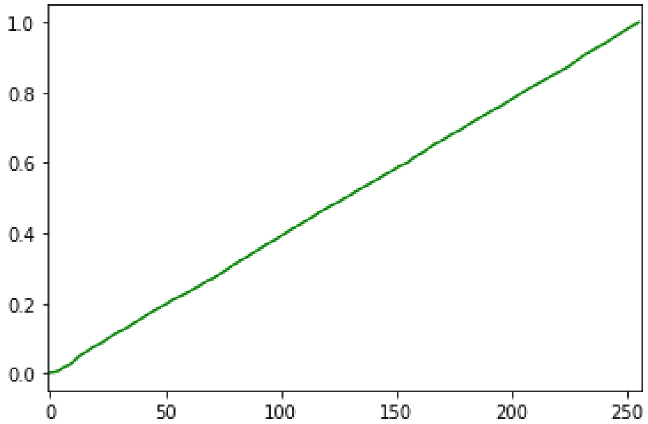
Cumulative Histogram



Ảnh kết quả



Density Histogram



Cumulative Histogram

Histogram equalization để tăng độ tương phản

Cho ảnh grayscale

Ảnh gốc

Ảnh kết quả

Cho ảnh màu

Ảnh gốc

Ảnh kết quả