WOMEN WHO CODE MANILA

Artificial Intelligence Study Group

Twitter: @wwcodemanila
FB: fb.com/wwcodemanila

#WWCodeManila
#AI
#StudyGroup

Issa Tingzon

Research Fellow

PCARI

New Member's Introduction

I am <name>

<your current profession>

<why did you join this study group?>

OUR MISSION

Inspiring women to excel in technology careers.

WOMEN WHO CODE MANILA

OUR VISION

A world where women are representative as technical executives, founders, VCs, board members and software engineers.

WOMEN WHO CODE MANILA

STUDY GROUP

Study groups are events where women can come together and help each other learn and understand a specific programming language, technology, or anything related to coding or engineering.

GUIDELINES

- If you have a question, just **ask**
- If you have an idea, **share it**
- **Make friends** and learn from your study groupmates
- **Do not** recruit or promote your business

OUTLINE

1. Introduction to Probability for ML
2. Baye's Theorem
3. Naïve Bayes Classifier

REVIEW

Probability for ML

Terminology

- **Probability** – deals with measuring likelihood around events

Terminology

- **Probability** – deals with measuring likelihood around events
- **Random Variable** –represents a domain of interest

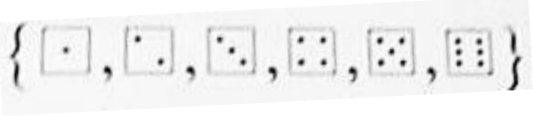
e.g. Modelling the probability of a die

Random variable X with **domain** equal to the possible outcomes {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}

Terminology

- **Probability** – deals with measuring likelihood around events
- **Random Variable** –represents a domain of interest

  e.g. Modelling the probability of a die

  Random variable X with **domain** equal to the possible outcomes 

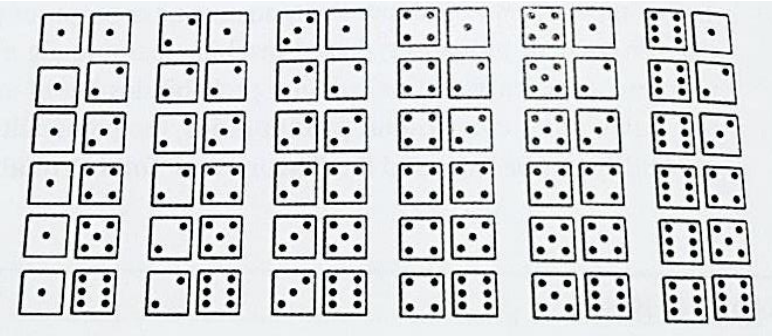
  Rolling 2 dice: 2 random variables $Dice_1$, $Dice_2$ each having this domain.

Terminology

- **Experiment** – rolling the two dice

Terminology

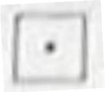
- **Experiment** – rolling the two dice
- **Sample Space** – set of all possible outcomes of the experiment



Sample space for the domain of two dice

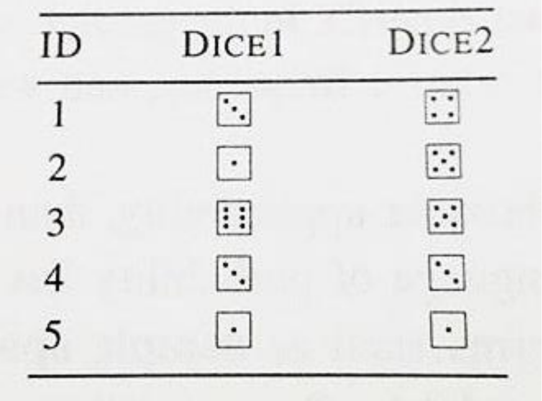
Terminology

- **Event** – an experiment whose outcome fixes the value of the random variable

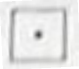
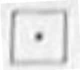
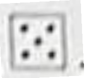
  e.g. $Dice_1 =$  , $Dice_2 =$ 

Terminology (Summary)

- Set of **random variables** → set of features in dataset
- **Sample space** → set of all combinations of assignments
- **Experiment** whose outcome has been recorded →
  row in the dataset

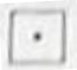
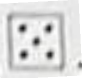


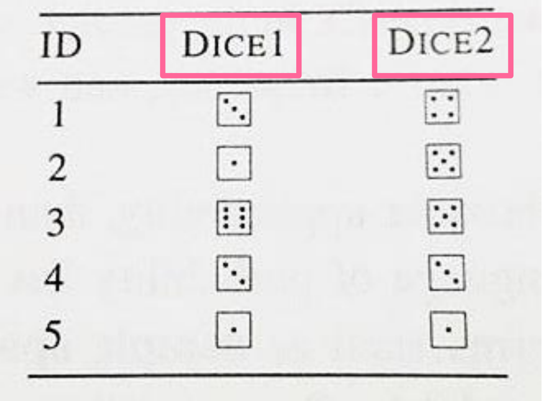
- **Experiment** whose outcome we do not know yet → prediction task
- **Event** → any subset of an experiment

e.g. $(Dice_1 =$  $)$, $(Dice_1 =$  $, Dice_2 =$  $)$

Terminology (Summary)

- Set of **random variables** → set of features in dataset
- **Sample space** → set of all combinations of assignments
- **Experiment** whose outcome has been recorded → row in the dataset
- **Experiment** whose outcome we do not know yet → prediction task
- **Event** → any subset of an experiment

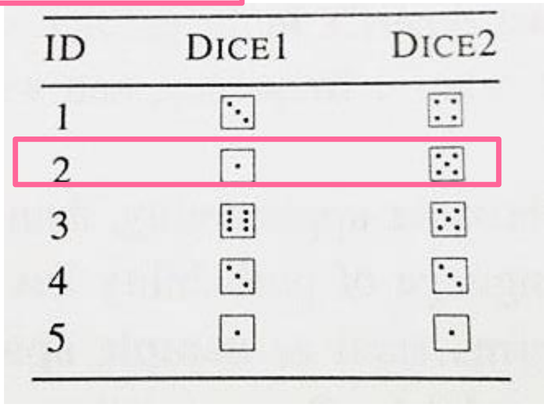
  e.g. ($Dice_1$= ) , ($Dice_1$=  , $Dice_2$ = )



Terminology (Summary)

- Set of **random variables** → set of features in dataset
- **Sample space** → set of all combinations of assignments
- **Experiment** whose outcome has been recorded → row in the dataset
- **Experiment** whose outcome we do not know yet → prediction task
- **Event** → any subset of an experiment

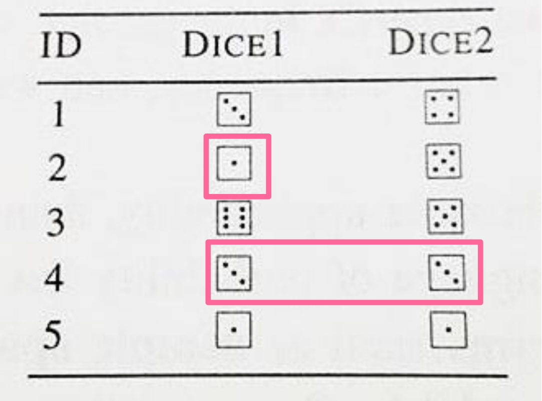
e.g. $(Dice_1=$ ⚀ $)$, $(Dice_1=$ ⚀ $, Dice_2 =$ ⚄ $)$



ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Terminology (Summary)

- Set of **random variables** → set of features in dataset
- **Sample space** → set of all combinations of assignments
- **Experiment** whose outcome has been recorded → row in the dataset



- **Experiment** whose outcome we do not know yet → prediction task
- $\boxed{\textbf{Event} \rightarrow \text{any subset of an experiment}}$

e.g. ($Dice_1$= ⚀) , ($Dice_1$= ⚀ , $Dice_2$ = ⚄)

Probability Functions

- Random variable can take on one or more values with a likelihood given by a *probability function* $P()$.

Probability Functions

- Random variable can take on one or more values with a likelihood given by a *probability function* $P()$.
- **Probability Function**
  - basic building block of probability theory
  - easy to create from the dataset
- $P()$ for an event is simply the **relative frequency** of the event in the dataset

Probability Functions

$$P() = \text{relative frequency} = \frac{how\ often\ the\ event\ happened}{how\ often\ it\ could\ have\ happened}$$

Probability Functions

$P()$ = relative frequency = $\dfrac{how\ often\ the\ event\ happened}{how\ often\ it\ could\ have\ happened}$

e.g. $P(Dice_1 =$ ⚀ $)$

ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Probability Functions

$P()$ = relative frequency = $\dfrac{how\ often\ the\ event\ happened}{how\ often\ it\ could\ have\ happened}$

e.g. $P(Dice_1 =$ ⚀ $) = \frac{2}{5} = 0.4$

ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Probability Functions

$P()$ = relative frequency = $\dfrac{how\ often\ the\ event\ happened}{how\ often\ it\ could\ have\ happened}$

e.g. $P(Dice_1 =$ ⚀ $) = \frac{2}{5} = 0.4$

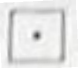
Joint probability – the probability of an event with more than one features

e.g. $P(Dice_1 =$ ⚂ $, Dice_2 =$ ⚃ $)$

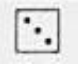
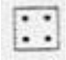
ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

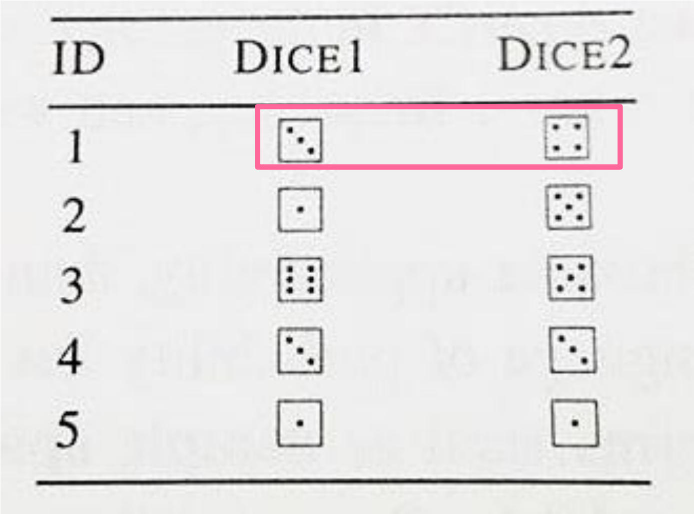
Probability Functions

$P()$ = relative frequency = $\dfrac{how\ often\ the\ event\ happened}{how\ often\ it\ could\ have\ happened}$

e.g. $P(Dice_1 =$  $) = \frac{2}{5} = 0.4$

Joint probability – the probability of an event with more than one features

e.g. $P(Dice_1 =$  $, Dice_2 =$  $) = \frac{1}{5} = 0.2$



ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Conditional and Unconditional Probabilities

- The types of probabilities we’ve calculated so far are known as **prior probabilities** or **unconditional probabilities**.

Conditional and Unconditional Probabilities

- The types of probabilities we've calculated so far are known as **prior probabilities** or **unconditional probabilities**.
- What about the probability of an event where one or more events are *known* to have happened?

Conditional and Unconditional Probabilities

- The types of probabilities we've calculated so far are known as **prior probabilities** or **unconditional probabilities**.
- What about the probability of an event where one or more events are *known* to have happened?
- **Posterior probability** or **conditional probability**

$P(X|Y)$ – probability that X will occur **given** Y has occured

read as *"given"*

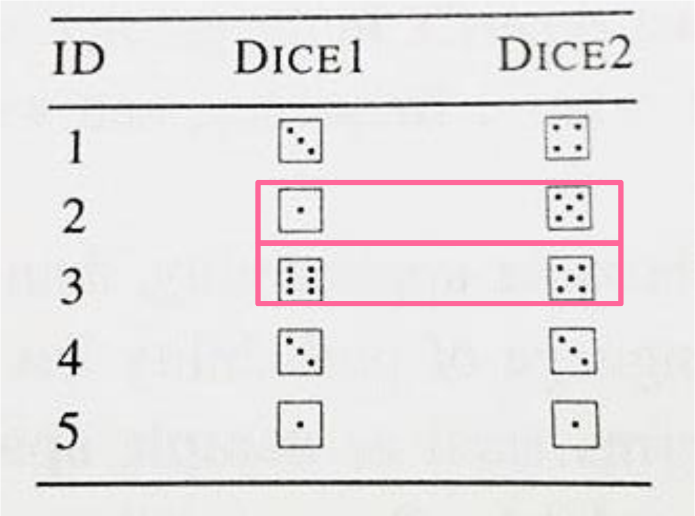
Aposteriori (conditional) Probability

e.g. $P(Dice_1 = ⚅ \mid Dice_2 = ⚄)$

ID	Dice1	Dice2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Aposteriori (conditional) Probability

e.g. $P(Dice_1 =$ ⚅ $|Dice_2 =$ ⚄ $) = \frac{\ }{2}$



ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

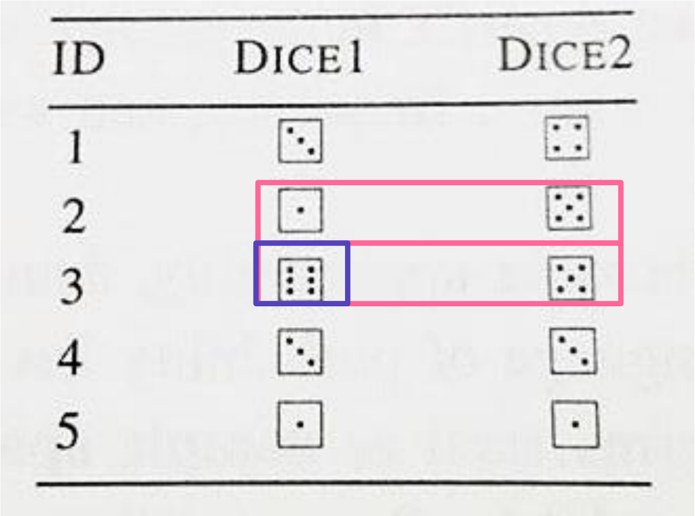
Aposteriori (conditional) Probability

e.g. $P(\boxed{Dice_1 = ⚅}|Dice_2 = ⚄) = \frac{1}{2}$

ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Aposteriori (conditional) Probability

e.g. $P(Dice_1 = \text{⚅} \mid Dice_2 = \text{⚄}) = \frac{1}{2} = 0.5$



ID	DICE1	DICE2
1	⚂	⚃
2	⚀	⚄
3	⚅	⚄
4	⚂	⚂
5	⚀	⚀

Quick Exercise 1

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

1. $P(H = true)$
2. $P(H = true, M = true)$
3. $P(M = true|H = true)$

Quick Exercise 1

ID	Headache	Fever	Vomiting	Meningitis
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

1. $P(H = true)$

Answer: $\frac{7}{10}$

2. $P(H = true, M = true)$

Answer: $\frac{2}{10}$

3. $P(M = true | H = true)$

Answer: $\frac{2}{7}$

Useful Property

- **Product Rule:** $P(X, Y) = P(X|Y)P(Y)$

  e.g. Check that:

  $$P(M = true, H = true) = P(M = true|H = true)P(H = true)$$

Useful Property

- **Product Rule:** $P(X,Y) = P(X|Y)P(Y)$

  e.g. Check that:

  $$P(M = true, H = true) = P(M = true|H = true)P(H = true)$$

- **Chain Rule:**

  $$P(A, B, \dots, Z) = P(A)P(B|A)P(C|B,A) \dots P(Z|Y, \dots, C, B, A)$$

Bayes' Theorem

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

- Proving this is pretty simple.

Bayesian Prediction

$$P(c|x_1,..,x_d) = \frac{P(x_1,\dots,x_d|c)P(c)}{P(x_1,\dots,x_d)}$$

$x = (x_1, \dots, x_d)$ represents an unknown experiment/instance

d – number of features

c – denotes a class label

Example

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

A patient comes in with a headache, vomiting, but no fever.

What is the probability that the patient has meningitis given these conditions?

Example

ID	Headache	Fever	Vomiting	Meningitis
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

A patient comes in with a headache, vomiting, but no fever.

What is the probability that the patient has meningitis given these conditions?

Answer: $\frac{1}{3}$

Exercise 3

ID	Headache	Fever	Vomiting	Meningitis
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

A patient comes in with a headache, vomiting, but no fever.

What is the probability that the patient **does not have** meningitis given these conditions?

Answer: $\frac{2}{3}$

Maximum Aposteriori Prediction (MAP)

$$c_{MAP} = \text{argmax}_{c \in Classes} \; P(c|x_1, .., x_d)$$

$$c_{MAP} = \text{argmax}_{c \in Classes} \; \frac{P(x_1, \dots, x_d|c)P(c)}{\cancel{P(x_1, \dots, x_d)}}$$

$$\boxed{c_{MAP} = \text{argmax}_{c \in Classes} P(x_1, \dots, x_d|c)P(c)}$$

Basically, the predicted class is the class label that returns the highest probability!

Maximum Aposteriori Prediction (MAP)

$$c_{MAP} = \text{argmax}_{c \in Classes} P(c|x_1, .., x_d)$$

$$c_{MAP} = \text{argmax}_{c \in Classes} \frac{P(x_1, \dots, x_d|c)P(c)}{\cancel{P(x_1, \dots, x_d)}}$$

$$\boxed{c_{MAP} = \text{argmax}_{c \in Classes} P(x_1, \dots, x_d|c)P(c)}$$

Basically, the predicted class is the class label that returns the highest probability!

For the patient with headache, vomiting, but no fever, what will be his predicted diagnosis in terms of meningitis?

Maximum Aposteriori Prediction (MAP)

$$c_{MAP} = \text{argmax}_{c \in Classes} P(c|x_1, .., x_d)$$

$$c_{MAP} = \text{argmax}_{c \in Classes} \frac{P(x_1, \ldots, x_d|c)P(c)}{\cancel{P(x_1, \ldots, x_d)}}$$

$$\boxed{c_{MAP} = \text{argmax}_{c \in Classes} P(x_1, \ldots, x_d|c)P(c)}$$

Basically, the predicted class is the class label that returns the highest probability!

For the patient with headache, vomiting, but no fever, what will be his predicted diagnosis in terms of meningitis?

Answer: Patient does not have meningitis.

Exercise 4

ID	Headache	Fever	Vomiting	Meningitis
11	true	true	false	false
37	false	true	false	false
42	true	false	true	false
49	true	false	true	false
54	false	true	false	true
57	true	false	true	false
73	true	false	true	false
75	true	false	true	true
89	false	true	false	false
92	true	false	true	true

A patient comes in with a headache, a fever, and vomiting.

What is the probability that the patient has meningitis given these conditions?

The Problem with Bayesian Prediction

- The problem is that our **dataset is not large enough** to be representative of the headache-fever-vomiting meningitis diagnosis scenario

The Problem with Bayesian Prediction

- The problem is that our **dataset is not large enough** to be representative of the headache-fever-vomiting meningitis diagnosis scenario
- In practice, it is impossible to collect a dataset that is big enough to cover all possible combinations.

The Problem with Bayesian Prediction

- The problem is that our **dataset is not large enough** to be representative of the headache-fever-vomiting meningitis diagnosis scenario
- In practice, it is impossible to collect a dataset that is big enough to cover all possible combinations.
- **Conditional independence** can help overcome these flaws.

Independence

- Two events are **independent** if one event has no effect on another event, ad vice-versa
- e.g. singing karaoke all night, and the event of rain in the morning

Independence

- Two events are **independent** if one event has no effect on another event, ad vice-versa
- e.g. singing karaoke all night, and the event of rain in the morning
- If two events X and Y are independent, then:

$$P(X|Y) = P(X)$$

$$P(X,Y) = P(X)P(Y)$$

Conditional Independence

- Full independence is quite rare
- More common: two or more events may be independent if we know that a third event has happened
- This is called **Conditional Independence**

Conditional Independence

- Full independence is quite rare
- More common: two or more events may be independent if we know that a third event has happened
- This is called **Conditional Independence**

e.g. $P(f|h)$ → knowing a patient has a headache increases the chance of him having a fever

$P(f|h,m)$ → if we already know the patient has meningitis, then knowing a patient has a headache **does not** increase the chance of him having a fever

Conditional Independence

- The information we get from knowing the patient has a headache is *contained within* the information that the patient has meningitis

Conditional Independence

- The information we get from knowing the patient has a headache is *contained within* the information that the patient has meningitis

- For two events X and Y that are conditionally independent given knowledge of a third event Z:

$$P(X|Y,Z) = P(X|Z)$$

$$P(X,Y|Z) = P(X|Z)P(Y|Z)$$

Conditional Independence

- Going back to this dilemma: $P(x_1, \ldots x_d|c)$
- **Chain rule**: $P(x_1, \ldots x_d|c) = \mathrm{P(x_1|c)P(x_2|x_1, c)} \cdots P(x_d|x_{d-1}, \ldots, x_2, x_{1,}, c)$

Conditional Independence

- Going back to this dilemma: $P(x_1, \dots x_d|c)$
- **Chain rule**: $P(x_1, \dots x_d|c) = \mathrm{P}(\mathrm{x}_1|\mathrm{c})\mathrm{P}(\mathrm{x}_2|\mathrm{x}_1, \mathrm{c}) \cdots P(x_d|x_{d-1}, \dots, x_2, x_{1,}, c)$
- Let's make a (strong) assumption that the features in an instance given a class label are conditionally independent.

$$P(x_1, \dots x_d|c) = \mathrm{P}(\mathrm{x}_1|\mathrm{c})\mathrm{P}(\mathrm{x}_2|\mathrm{c}) \cdots P(x_d|c)$$

$$P(x_1, \dots x_d|c) = \prod_{i=1}^{d} P(x_i|c)$$

Naïve Bayes Classifier

- Returns a **MAP prediction**, where the posterior (conditional) probabilities are computed under the **strong assumption of conditional independence** between features given the class label.

$$c_{MAP} = \mathrm{argmax}_{c \in Classes} P(c) \prod_{i=1}^{d} P(x_i | c)$$

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

1. Count the statistics.
2. Will John play tennis if the outlook is sunny, humidity is high, and wind is strong?
3. What about when it’s rainy, highly humid, and not windy?

Naïve Bayes Classifier

$$c_{MAP} = \text{argmax}_{c \in Classes} P(c) \prod_{i=1}^{d} P(x_i|c)$$

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Will a Red domestic SUV be likely to be stolen or not?

References

Decision Tree Lecture by Victor Lavrenko (Youtube)

Random Forests Video by Siraj Raval

T.I.L.

SHARE IT!

In front!

On Twitter: @wwcodemanila

Or FB: fb.com/wwcodemanila

Don’t forget to tag WWCodeManila so we can retweet or share it.

Feedback Form

https://goo.gl/YzSqcS

Please don't rate the event on meetup.
Not helpful. It is best to just tell your concerns via the feedback form. We are a building a community not a Yelp restaurant.

THANK YOU :)