MACHINE LEARNING ALGORITHM Unit-III

Dr. Gopal Sakarkar
Department of AI, GHRCE, Nagpur

YouTube Channel on Machine Learning Algorithms
https://tinyurl.com/GopalMachineLearningAlgorithms

Unit-III

Probability and Bayes learning

Probability

- *Probability* is the study of randomness and uncertainty.
- A *random experiment* is a process whose outcome is uncertain.

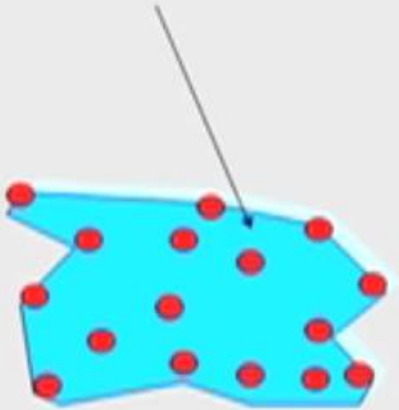
  Examples:
  - Tossing a coin once or several times
  - Tossing a die
  - Tossing a coin until one gets Heads
  - ...

Probability

Events and Sample Spaces

Sample Space

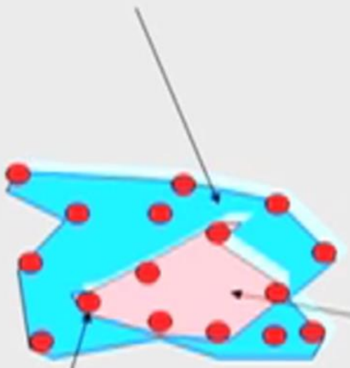
The sample space is the set of all possible outcomes.



Probability

Events and Sample Spaces

Sample Space
The sample space is the set of all possible outcomes.



Simple Events
The individual outcomes are called simple events.

Event
An event is any collection of one or more simple events

Probability

Sample Space

- Sample space Ω : the set of all the possible outcomes of the experiment
  - If the experiment is a roll of a six-sided die, then the natural sample space is {1, 2, 3, 4, 5, 6}
  - Suppose the experiment consists of tossing a coin three times.
    $\Omega = \{(hhh, hht, hth, htt, thh, tht, tth, ttt\}$
  - the experiment is the number of customers that arrive at a service desk during a fixed time period, the sample space should be the set of nonnegative integers: $\Omega = Z^{+} = \{0, 1, 2, 3, \ldots\}$

Probability

Events

- Events are subsets of the sample space
  - A= {the outcome that the die is even} ={2,4,6}
  - B = {exactly two tosses come out tails}=(htt, tht, tth}
  - C = {at least two heads} = {hhh, hht, hth, thh}

Probability

Discrete Random Variables

- Random variables (RVs) which may take on only a countable number of distinct values
  - e.g., the sum of the value of two dies
- X is a RV with arity *k* if it can take on exactly one value out of k values,
  - e.g., the possible values that X can take on are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Calculating Probability

Probability $P(A)$ for an event A which can occur in m ways out of n equally likely opportunities is given by:

$$P(A) = \frac{m}{n}$$

Calculating Probability

Probability P(A) for an event A which can occur in *m* ways out of *n* equally likely opportunities is given by:

$$P(A) = \frac{m}{n}$$

Probability ranges between 0 and 1, where 0 implies an impossible event and 1 implies an event which is definite or 100% certain.

Complementary Rule

SAMPLE SPACE

sample space (also called sample description space or possibility space) of an experiment or random trial is the set of all possible outcomes or results of that experiment.

Complementary Rule

SAMPLE SPACE

Complementary Rule

SAMPLE SPACE

Complement

- The Complement of an event is all the other outcomes (not the ones we want).
- Together the Event and its Complement make all possible outcomes.

$$P(E) + P(E') = 1$$

Example for Complement

- According to the American Veterinary Medical Association, 31.6% of American households own a dog. What is the probability that a randomly selected household does not own a dog?
  - E= Own a dog $P(\overline{E}) = 1 - P(E)$
  - P(E) =31.6% $P(\overline{E}) = 68.4\%$

Independent Event

- Independent events are events such that the outcome of one event does not affect the outcome of the second, and vice versa.
  - For Ex: Event A: It rained on Tuesday.
  - Event B: My chair broke at work.
  - These two events are unrelated. Probability of one event is not going to affect another event.

Mutually Exclusive Vs Independent

- if A and B are mutually exclusive, they cannot be independent. Because it make other event probability to be zero. (It affecting other event probability).

Multiplication rule for Independent event

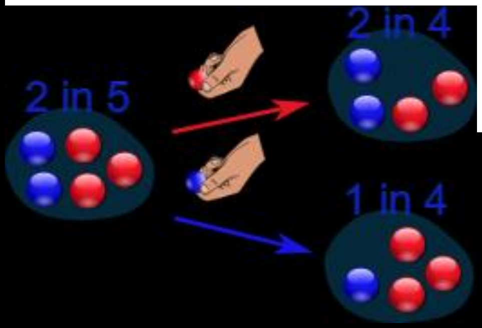
$$P(A \text{ and } B) = P(A \cap B) = P(A) * P(B)$$

Example : Suppose we roll one die followed by another and want to find the probability of rolling a 4 on the first die and rolling an even number on the second die.

$$P(A \cap B) = 1/12$$

Conditional Probability

- What is the probability of drawing two blue marbles without replacement one by one?
  - Event A = Drawing blue marble first.
  - Event B = Drawing blue marble second.
- P(A) = 2/5 (Probability of drawing blue marble first).
- P(B|A) = ¼ (Event A has happened, what is the probability of Event B).



$P(A \cap B) = 1/$

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

Examples for Conditional Probability

- 70% of your friends like Chocolate, and 35% like Chocolate AND like Strawberry. What percent of those who like Chocolate also like Strawberry?
- Two cards are selected without replacement, from a standard deck. Find the probability of selecting a king and then selecting a queen.

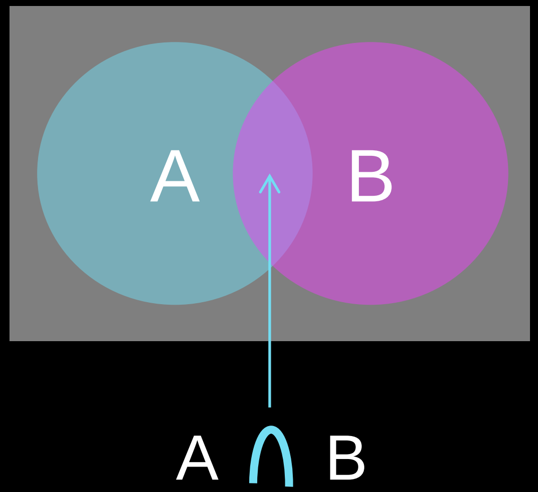
Bayes’ Theorem

- The Bayes’ Theorem was developed and named for Thomas Bayes (1702 – 1761).
- It can be seen as a way of understanding how the probability that a theory is true is affected by a new piece of evidence.

Bayes' Theorem

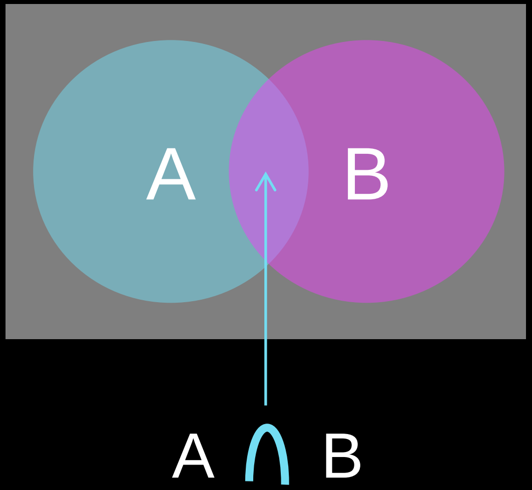
- Bayes's theorem, Bayes's law or Bayes's rule : Describes the probability of an event, based on prior knowledge of conditions that might be related to the event

Conditional Probability



Where, Gray rectangle space is sample space.
It is a collection of possible outcomes.
A and B are events, they are overlapping .
The portion common to A and B is A Intersection B.

Conditional Probability



$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

What is probability of occurrence of A where B is already occurred .
Its read out like Probability of occurrence of A given that B is already occurred.

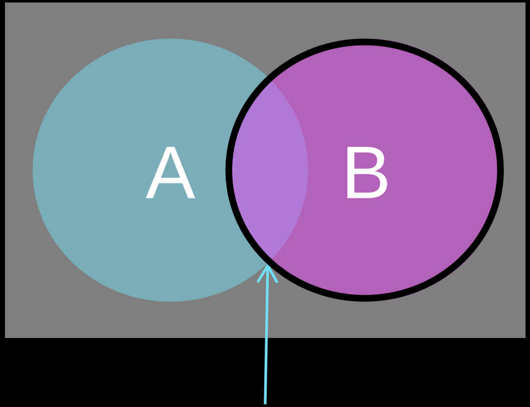
Conditional Probability

Probability of A and B

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

Probability of A given B

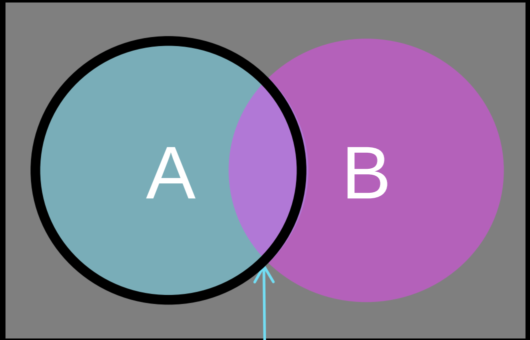
Probability of B



$A \cap B$

Here assume that, an entire sample space has shrank to event B which is already occurred.

Conditional Probability



A ∩ B

Probability of A and B

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

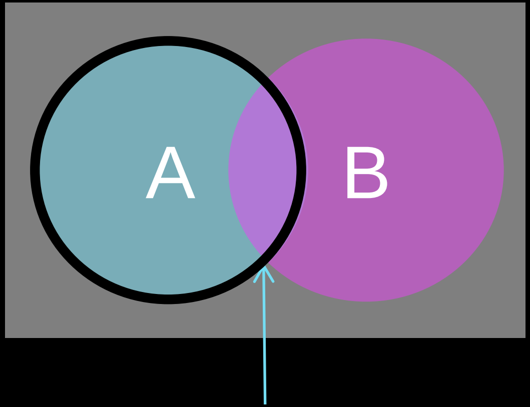
Probability of A given B

Probability of B

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

Here, probability of occurrence B given that event A is already occurred . Assuming that sample space shrank to A.

Conditional Probability



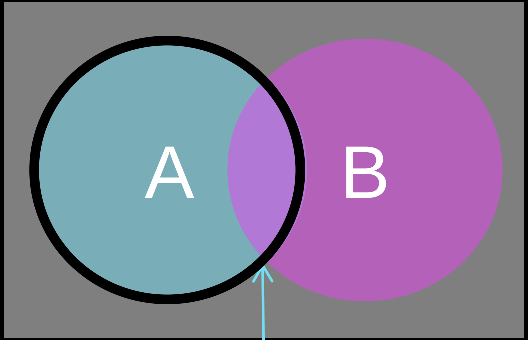
A ∩ B

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

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

$$P(A \cap B) = P(A \mid B) * P(B) = P(B \mid A) * P(A) \text{ - - (I)}$$

Conditional Probability



A ∩ B

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

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

$$\boxed{P(A \cap B) = P(A \mid B) * P(B) = P(B \mid A) * P(A) \quad - - \text{(I)}}$$

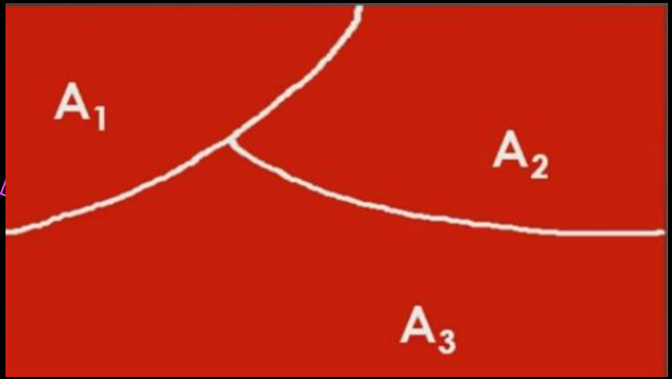
$$P(A \mid B) * P(B) = P(B \mid A) * P(A)$$

$$\boxed{P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} \quad - - \text{(II)}}$$

Sample Space & Intersection of Events

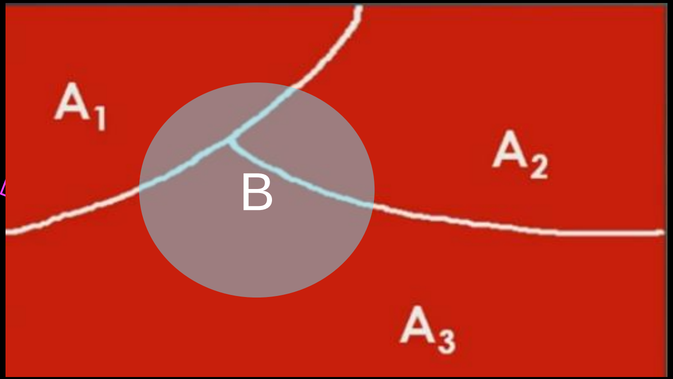
- Let us consider three events, A1, A2 and A3.
- They are Mutually Exclusive and Collectively Exhaustive.
- Mutual exclusive means occurrence of one event neglects the possibility of other events.
- e.g. when you toss a coin, you either get Head or Tail , not get Head and Tail at same time.

- Collectively Exhaustive means when rolling a six-sided die, the outcomes 1, 2, 3, 4, 5, and 6 are collectively exhaustive, because they encompass the entire range of possible outcomes.

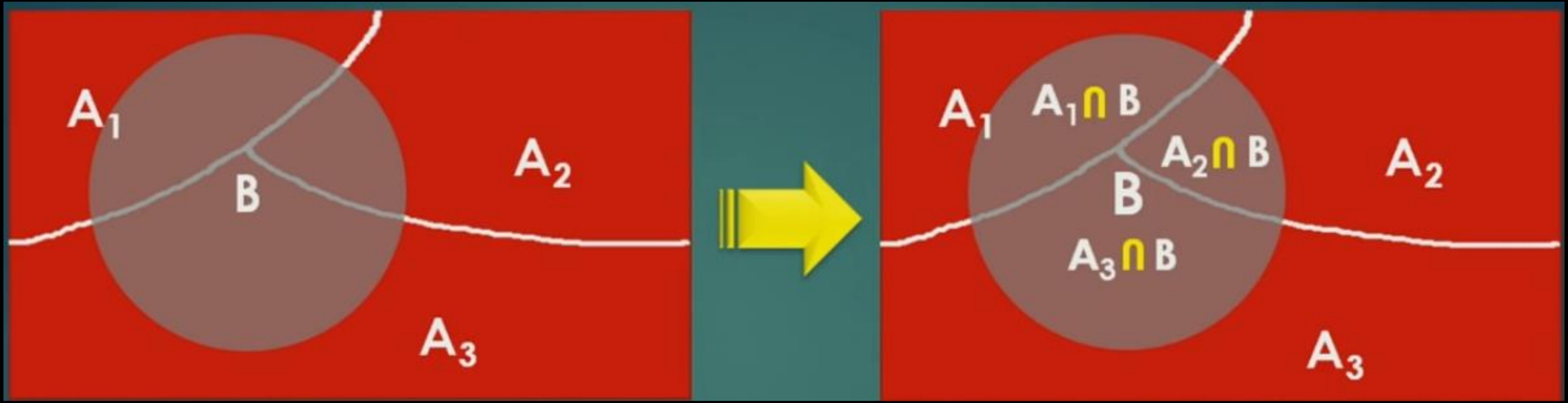


Sample Space & Intersection of Events

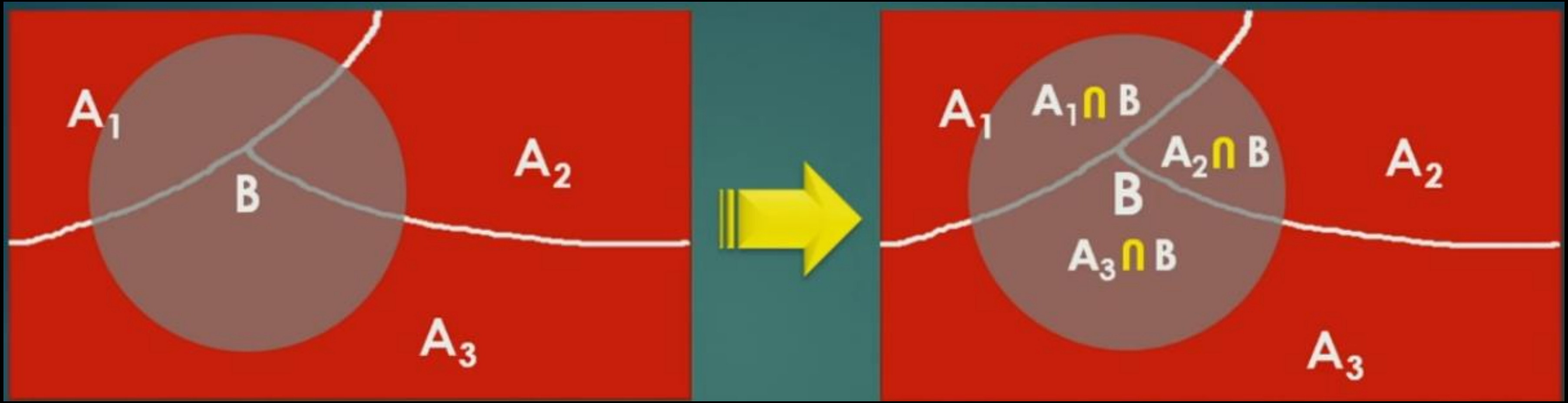
- Let B is occurred , which has some common A1, A2 and A3



Sample Space & Intersection of Events

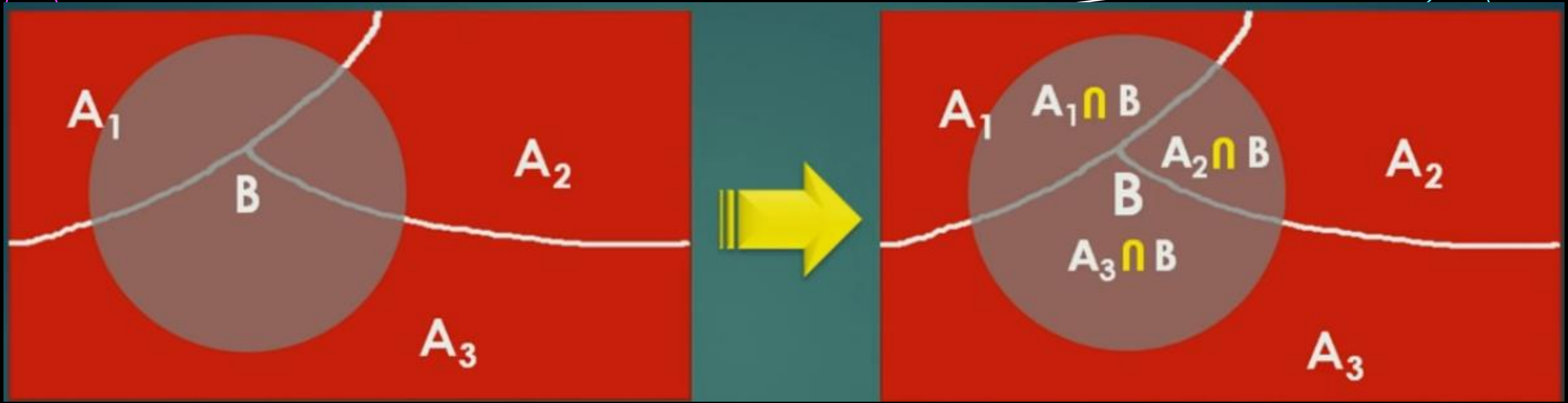


Sample Space & Intersection of Events



$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

Sample Space & Intersection of Events

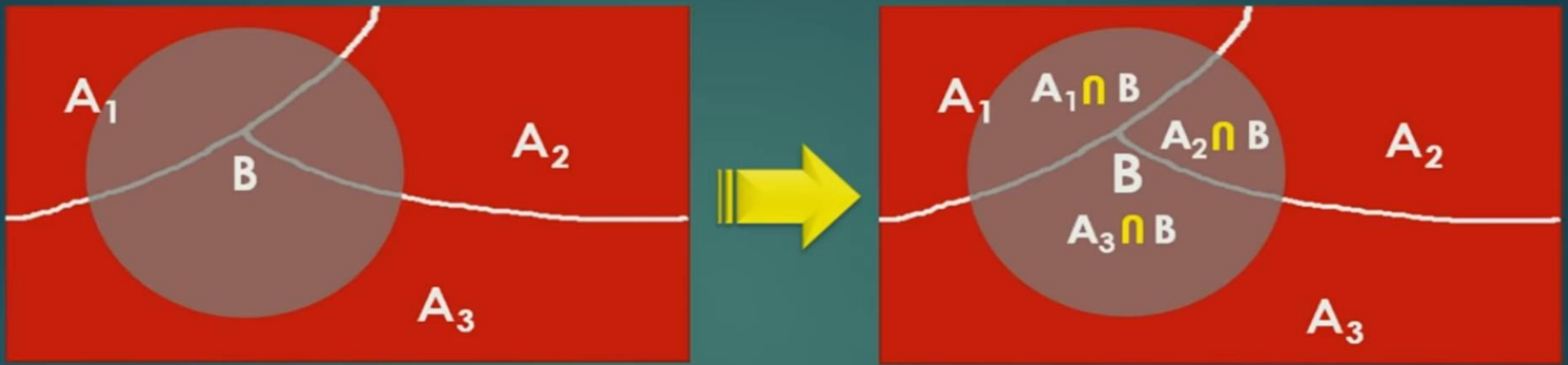


$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

$$P(B) = P(B \mid A_1)*P(A_1) + P(B \mid A_2)*P(A_2) + P(B \mid A_3)*P(A_3)$$

From equation no. I

Sample Space & Intersection of Events



$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

$$P(B) = P(B \mid A_1)*P(A_1) + P(B \mid A_2)*P(A_2) + P(B \mid A_3)*P(A_3)$$

Its generalized equation as :

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)*P(A_i) \quad \text{(iii)}$$

Putting it all together

$$P(A_i \mid B) = \frac{P(B \mid A_i) * P(A_i)}{P(B)} \quad \text{- - - -(ii)}$$

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) * P(A_i) \quad \text{- - - -(iii)}$$

Putting it all together

$$P(A_i \mid B) = \frac{P(B \mid A_i)*P(A_i)}{P(B)} \quad \text{---(ii)}$$

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)*P(A_i) \quad \text{---(iii)}$$

$$\boxed{P(A_i \mid B) = \frac{P(B \mid A_i)*P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)*P(A_i)}}$$

In order to manage the Credit Risk, a bank regularly rates each of its borrowers as A_1 or A_2 or A_3, based on their Credit history. A_1 implies lowest risk and A_3 implies highest risk. Risk means the chance that a borrower might fail to payback the loan amount.

Based on historical data, on an average, 30% customers are rated A_1, 60% are rated A_2, and 10% are rated A_3. It was found that 1% of the customers who were rated A_1, 10% of the customers who were rated A_2, and 18% of the customers who were rated A_3, eventually became defaulters(failed to payback).

If you randomly pickup a customer from defaulter's pool, what is the probability that he had received an A_1 rating?

Bayes' Theorem

Step 1 - What is being asked?

*"If you randomly pickup a customer from **defaulter's pool**, what is the probability that he had received an **A_1 rating**?"*

P(**Rating A_1** | **Defautler**) = ?

Note: Flip of what is being asked i.e. **P(Defautler | Rating A_1)** will always be given in such problems.

*"It was found that **1% of the customers who were rated A_1**, 10% of the customers who were rated A_2, and 18% of the customers who were rated A_3, **eventually became defaulters** (failed to payback)."*

Bayes' Theorem

Step 2 - What is given?

"Based on historical data, on an average, 30% customers are rated A_1, 60% are rated A_2, and 10% are rated A_3."

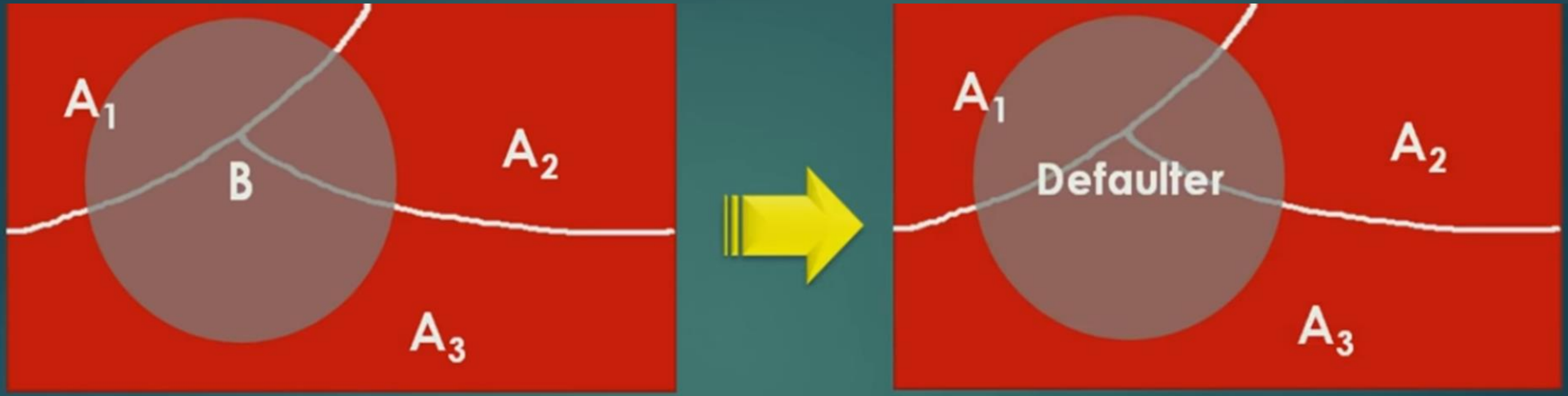
P(Defautler | Rating A_1) = 1% or 0.01
P(Defautler | Rating A_2) = 10% or 0.10
P(Defautler | Rating A_3) = 18% or 0.18

P(Rating A_1) = 30% or 0.30
P(Rating A_2) = 10% or 0.60
P(Rating A_3) = 18% or 0.10

Note: P(Rating A_1) + P(Rating A_2) + P(Rating A_3) = 0.30 + 0.60 +0.10 = 1

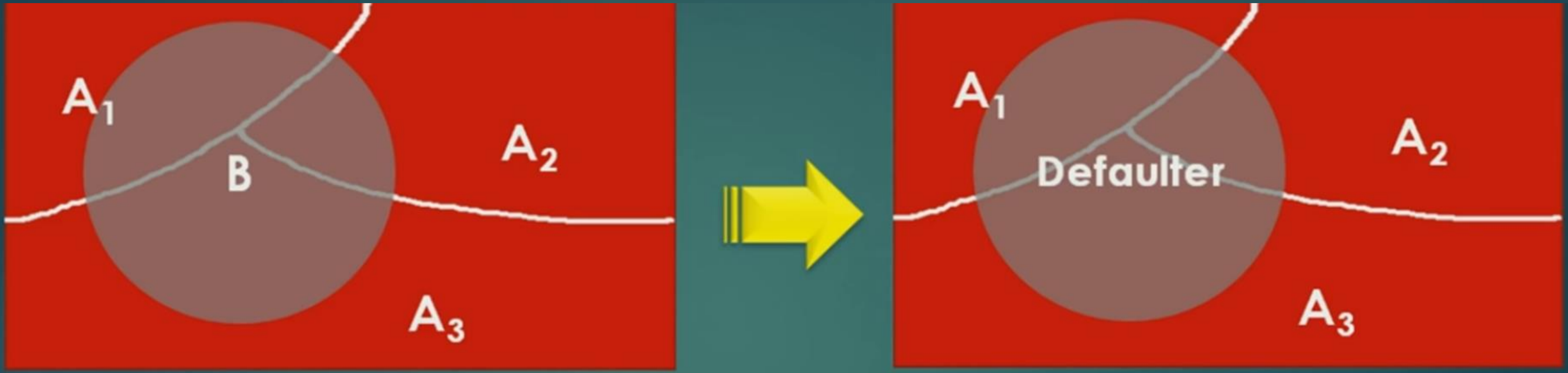
Bayes' Theorem

Step 2 - What is given?



Bayes' Theorem

Step 2 - What is given?



$$P(A_i \mid B) = \frac{P(B \mid A_i)*P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)*P(A_i)}$$

Bayes' Theorem

Step 3 - Put the values

$$P(A_i \mid B) = \frac{P(B \mid A_i)*P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)*P(A_i)}$$

Numerator:

$P(\text{Defautler} \mid \text{Rating } A_1)*P(\text{Rating } A_1) = 0.01*0.30 = 0.003$ From Eq: 1

Denominator:

$$\sum_{i=1}^{n} P(B \mid A_i)*P(A_i) = P(B \mid A_1)*P(A_1) + P(B \mid A_2)*P(A_2) + P(B \mid A_3)*P(A_3)$$

$$= 0.01*0.30 + 0.10*0.60 + 0.18*.10$$

$$= 0.081$$

$$P(\text{Rating } A_1 \mid \text{Defautler}) = \frac{0.003}{0.081} = 0.0370 \text{ or } 3.7\%$$

Bayes' Theorem

Step 1 - What is being asked?

*"If you randomly pickup a customer from **defaulter's pool**, what is the probability that he had received an **A_1 rating**?"*

Answer is 3.7 %

Bayes' Theorem Example

- In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts.
- If a patient is an addict, what is the probability that they will be prescribed pain pills?

Bayes' Theorem Example

- Step 1: Figure out what your event "A" is from the question. That information is in the italicized part of this particular question. The event that happens first (A) is being prescribed pain pills. That's given as 10%.
- Step 2: Figure out what your event "B" is from the question. That information is also in the italicized part of this particular question. Event B is being an addict. That's given as 5%.
- Step 3: Figure out what the probability of event B (Step 2) given event A (Step 1). In other words, find what (B|A) is. We want to know "Given that people are prescribed pain pills, what's the probability they are an addict?" That is given in the question as 8%, or .8.
- Step 4: Insert your answers from Steps 1, 2 and 3 into the formula and solve. $P(A|B) = P(B|A) * P(A) / P(B) = (0.08 * 0.1)/0.05 = 0.16$
- The probability of an addict being prescribed pain pills is 0.16 (16%).

Bayes' Theorem Example

- 1% of people have a certain genetic defect.
- 90% of tests for the gene detect the defect (true positives).
- 9.6% of the tests are false positives.
- If a person gets a positive test result, what are the odds they actually have the genetic defect?

- The first step into solving Bayes' theorem problems is to assign letters to events:

- A = chance of having the faulty gene. That was given in the question as 1%. That also means the probability of not having the gene (~A) is 99%.
- X = A positive test result.
- So:

- P(A|X) = Probability of having the gene given a positive test result.
- P(X|A) = Chance of a positive test result given that the person actually has the gene. That was given in the question as 90%.
- p(X|~A) = Chance of a positive test if the person doesn't have the gene. That was given in the question as 9.6%
- Now we have all of the information we need to put into the equation:
- P(A|X) = (.9 * .01) / (.9 * .01 + .096 * .99) = 0.0865 (8.65%).

- The probability of having the faulty gene on the test is 8.65%

Bayes' Theorem Example

- A doctor is called to see a sick child. The doctor has prior information that 90% of sick children in that neighborhood have the flu, while the other 10% are sick with 1 measles.
- Let F stand for an event of a child being sick with flu and M stand for an event of a child being sick with measles. Assume for simplicity that $F \cup M = \Omega$, i.e., that there no other maladies in that neighborhood.
- A well-known symptom of measles is a rash (the event of having which we denote R).
- Assume that the probability of having a rash if one has measles is $P(R \mid M) = 0.95$.
- However, occasionally children with flu also develop rash, and the probability of having a rash if one has flu is $P(R \mid F) = 0.08$.
- Upon examining the child, the doctor finds a rash. What is the probability that the child has measles?

Probability from Tree

- G= Economic Grow
- S = Economic Slow
- U = Stock up
- D = Stock Down
- P(G) = Probability of Economic Grows (70%)
- P(U/G) = Probability of stock improves (up) given that economy is growing. (80%) (Conditional probability : What is the probability of stock up with condition on economy is growing?

Probability

- What is the probability of economy grows **and** stock up?

$$P(G \cap U) = P(U/G) \times P(G)$$
$$= 0.7 \times 0.8 = 0.56$$

- What is the probability that economy grows given that stock went up?

$$P(G/U) = \text{Apply Bayes' Theorem}$$

Bayes' Theorem

$$P(G|U) = \frac{P(U|G) * P(G)}{P(U)}$$

$$P(G|U) = \frac{P(U|G) * P(G)}{P(U|G) * P(G) + P(U|S) * P(S)}$$

$$= 86\%$$

- P(G) = 70% (Unconditional Probability)
- By giving addition of new information that stock went up, unconditional probability becomes conditional probability P(G|U) = 86%
- This is called Bayes' Theorem

Exercise - 1

- 1% of the population has X disease. A screening test accurately detects the disease for 90% of people with it. The test also indicates the disease for 15% of the people without it (the false positives). Suppose a person screened for the disease tests positive. What is the probability they have it?

Given:

$P(D) = .01$

$P(T|D) = 0.9$

$P(T|\grave{D}) = 0.15$

Find:

$P(D|T) = ?$

Exercise - 2

- Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

W = Correctly predicted by Weathermen.

R = Rain

1. P(R) = 5/365
2. P (W|R) = 90%
3. P ($\grave{W}|\grave{R}$) = 10%
4. P(R|W) = ?

Exercise

- A single 6-sided die is rolled. What is the probability of rolling a 2 or a 5?
- A glass jar contains 1 red, 3 green, 2 blue, and 4 yellow marbles. If a single marble is chosen at random from the jar, what is the probability that it is yellow or green?
- A single card is chosen at random from a standard deck of 52 playing cards. What is the probability of choosing a king or a club?
- A number from 1 to 10 is chosen at random. What is the probability of choosing a 5 or an even number?

Bayes Learning/ Bayes Theorem :Exercise

To find the probability of selected card will be King card , given that list of Face cards.

Given : P(King/Face)

$$P(King/Face)= \frac{P(Face/King) \times P(King)}{P(Face)}$$

Putting Values

$$= \frac{1 \times 4/52}{12/52}$$ –(only 4 King cards out of 52)

Whatever a card you selected from king card it would be face card

Bayes Learning/ Bayes Theorem :Exercise

To find the probability of selected card will be King card , given that list of Face cards.

Given : P(King/Face)

$$P(King/Face) = \frac{P(Face/King) \times P(King)}{P(Face)}$$

Putting Values

$$= \frac{1 \times 4/52}{12/52}$$

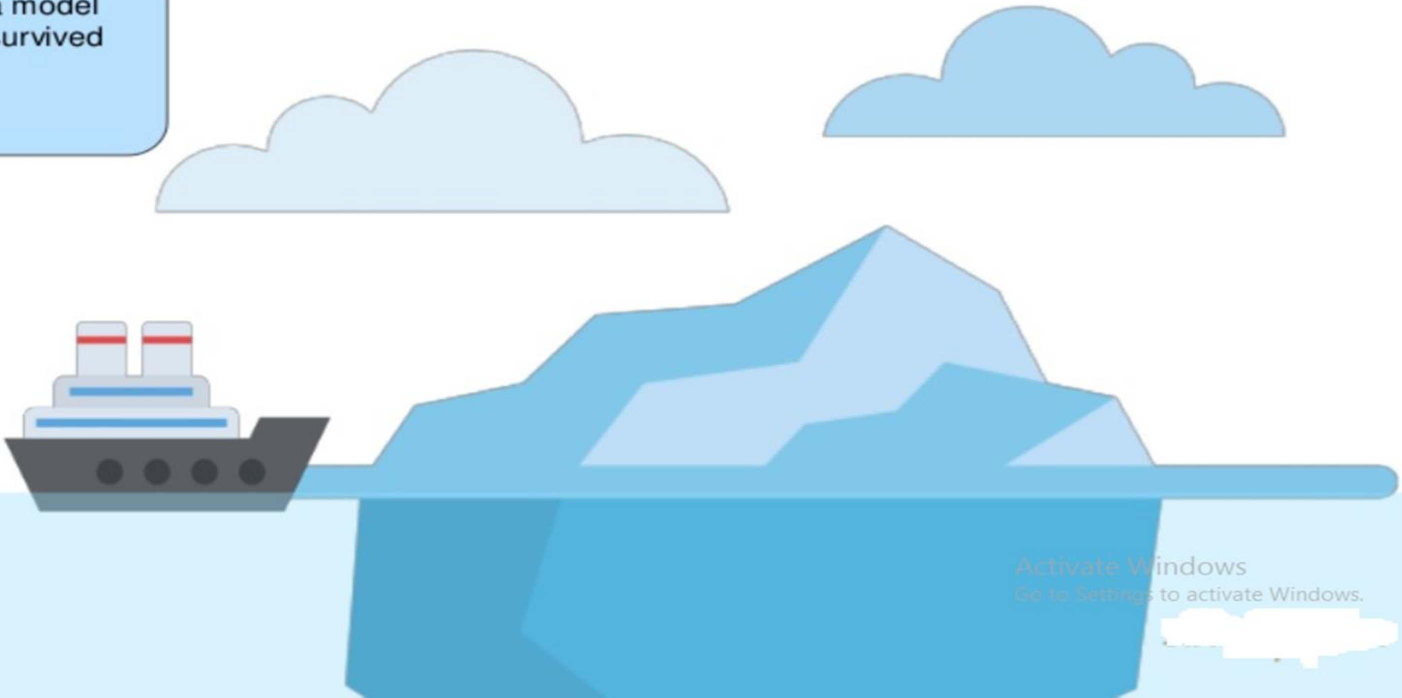
$$= \frac{1 \times 1/13}{3/13}$$

$$= 1/3$$

Logistic Regression

Surviving the Titanic

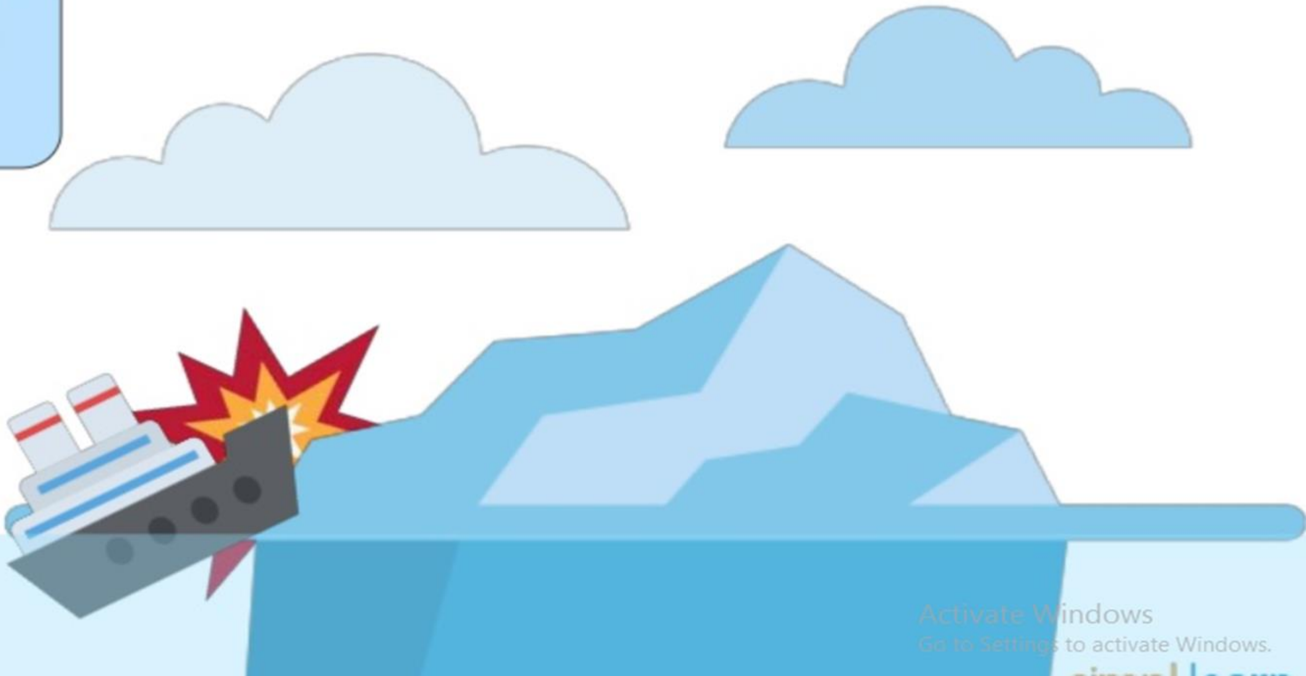
Suppose, you have to build a model to predict how many people survived the Titanic shipwreck



Logistic Regression

Surviving the Titanic

Suppose, you have to build a model to predict how many people survived the Titanic shipwreck



Logistic Regression

Surviving the Titanic

Teaching the model with the passenger dataset

Dropping the non-essential components of the dataset

Determining the survival of passengers and evaluating the model

Logistic Regression

Agenda

- What is Supervised Learning?
- What is Classification? What are some of its solutions?
- What is Logistic Regression?
- Comparing Linear and Logistic Regression
- Logistic Regression applications
- Use Case – Predicting the number in an image

Logistic Regression

Logistic Regression

What is Supervised Learning?

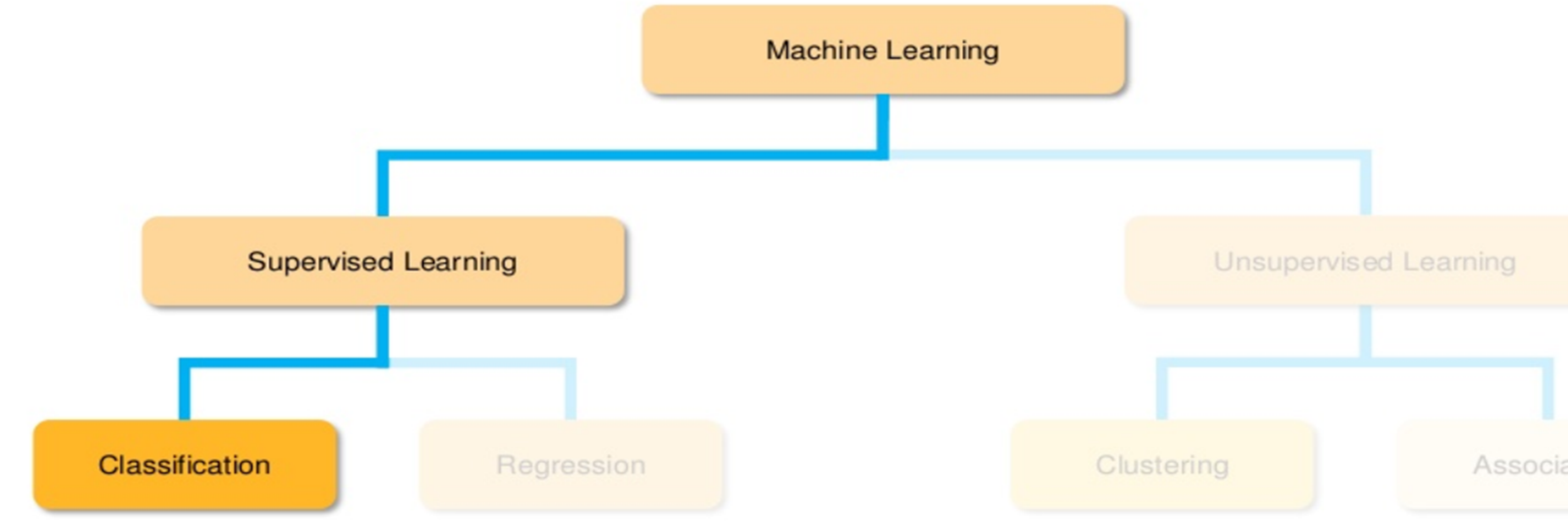
Teacher teaches child

Child recognizes an apple when she sees it again

A model is able to make predictions based on past data

Logistic Regression

Where does Logistic Regression fit it?

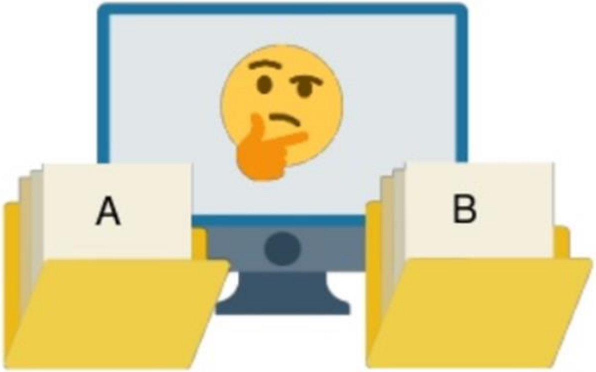


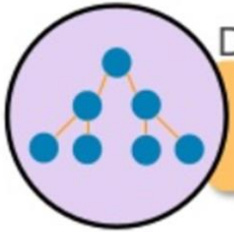
The systems predicts future outcomes based on training from past input

Logistic Regression

Logistic Regression

A few Classification Solutions



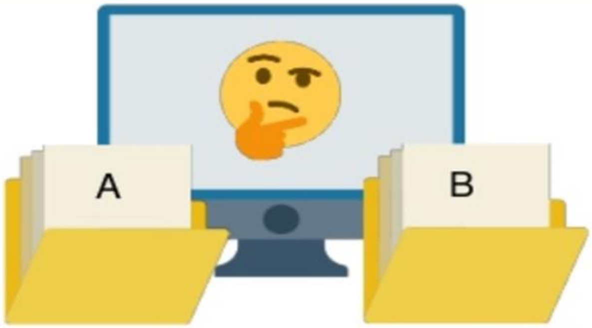


Decision Trees

We take decisions using a tree structure. Each branch node represents a choice, and leaf node represents a decision

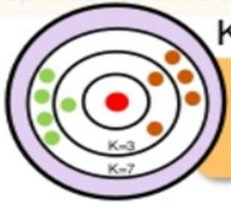
Logistic Regression

A few Classification Solutions



Decision Trees

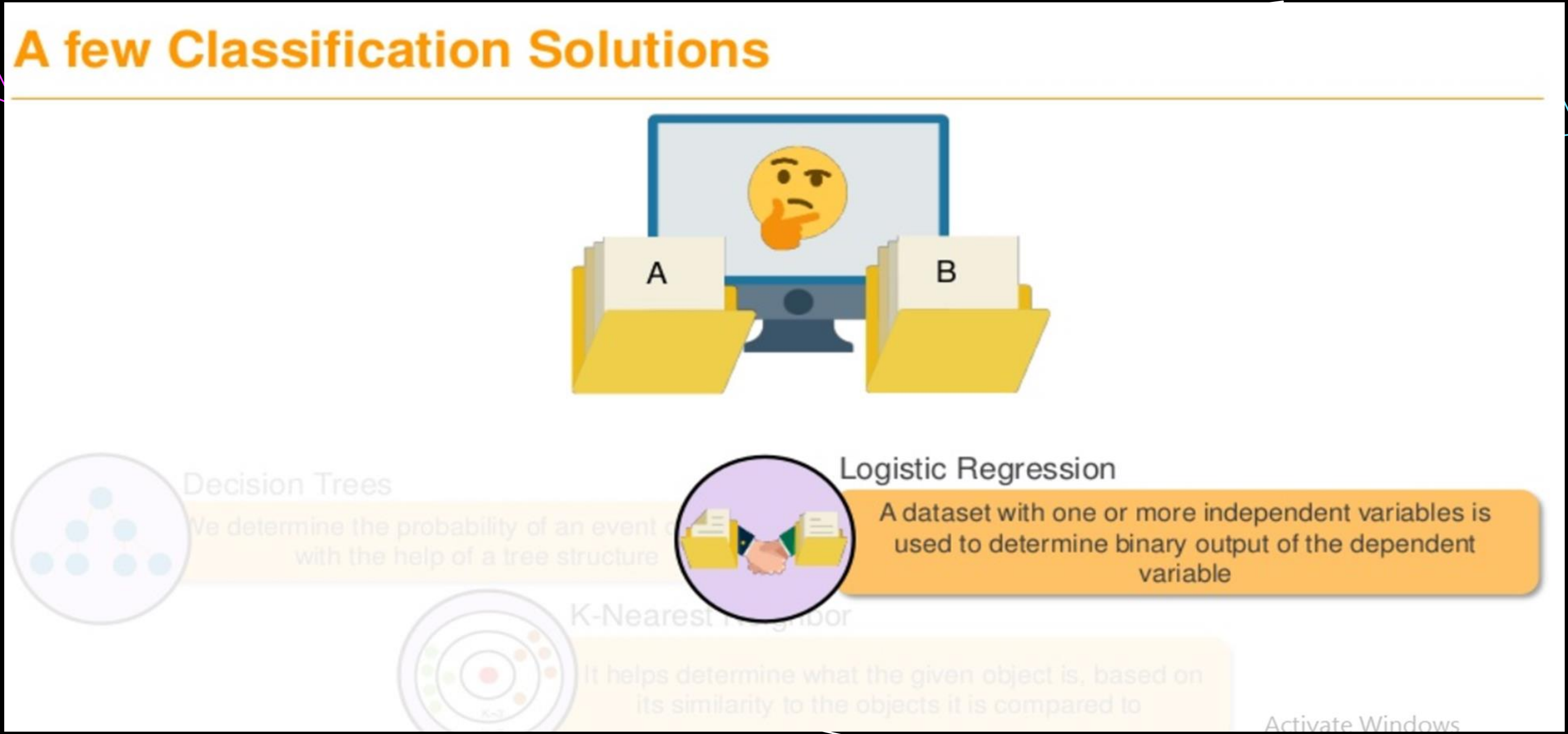
We take decisions using a tree structure. Each branch node represents a choice, and leaf node represents a decision



K-Nearest Neighbor

It helps determine what the given object is, based on its similarity to the objects it is compared to

Logistic Regression



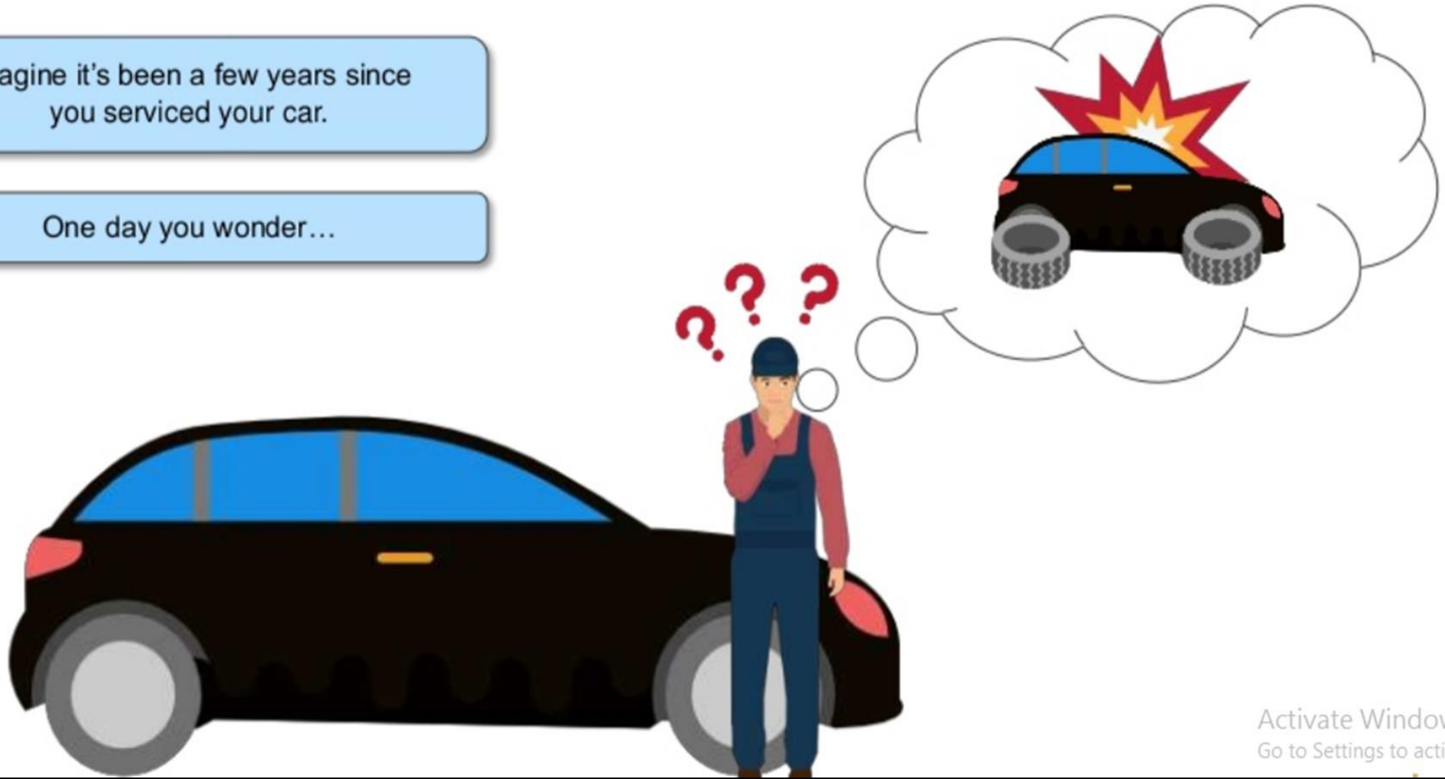
Logistic Regression

Logistic Regression

What is Logistic Regression?

Imagine it's been a few years since you serviced your car.

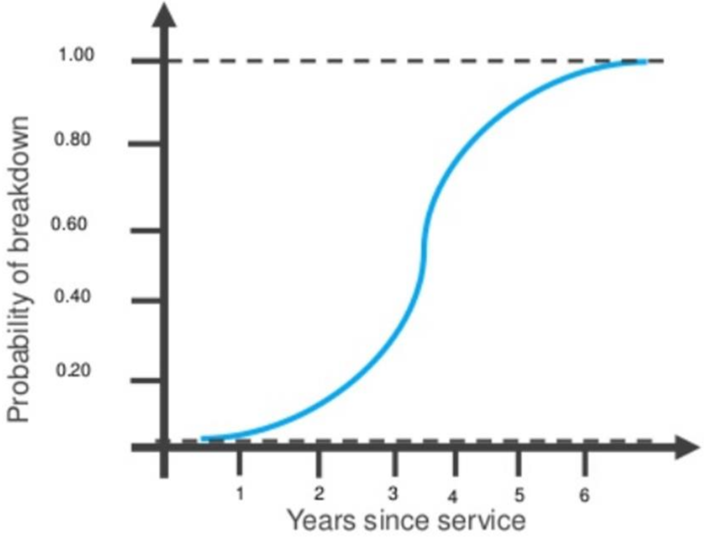
One day you wonder…



Logistic Regression

What is Logistic Regression?

You provide years since last service

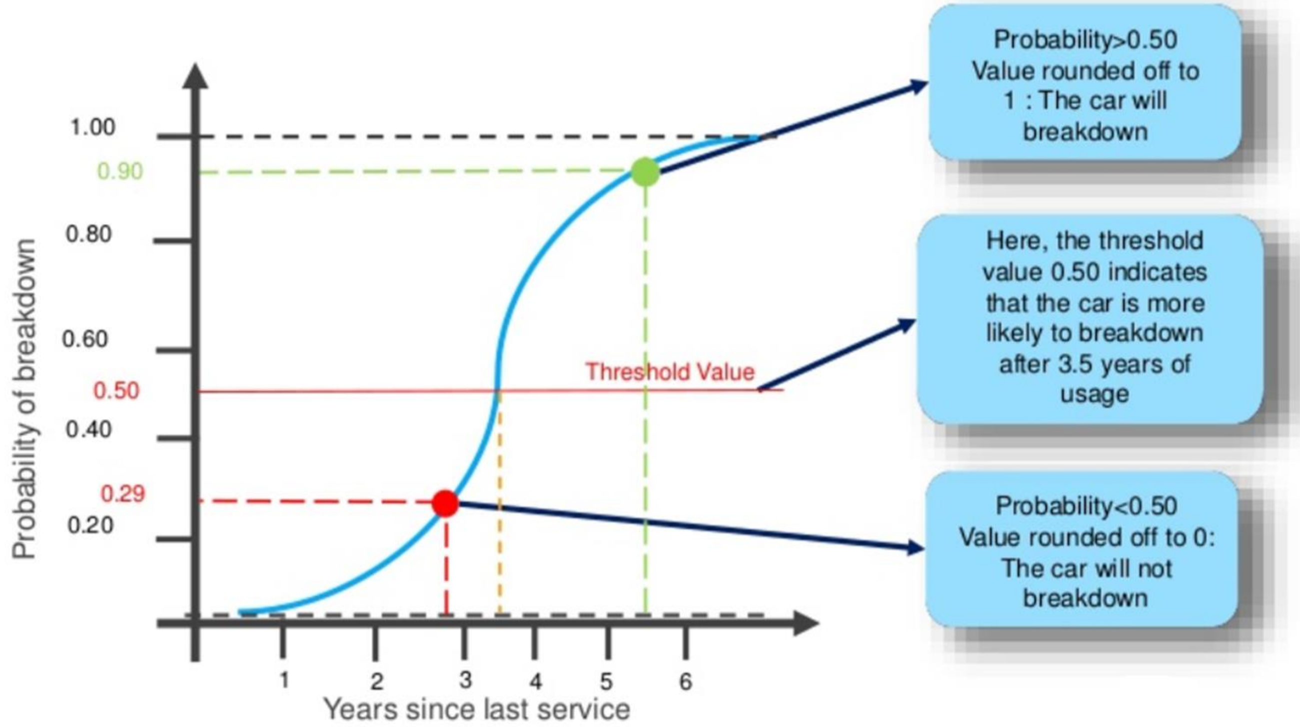


Regression model created based on other users' experience

It is a classification algorithm, used to predict binary outcomes for a given set of independent variables. The dependent variable's outcome is discrete.

Logistic Regression

What is Logistic Regression?

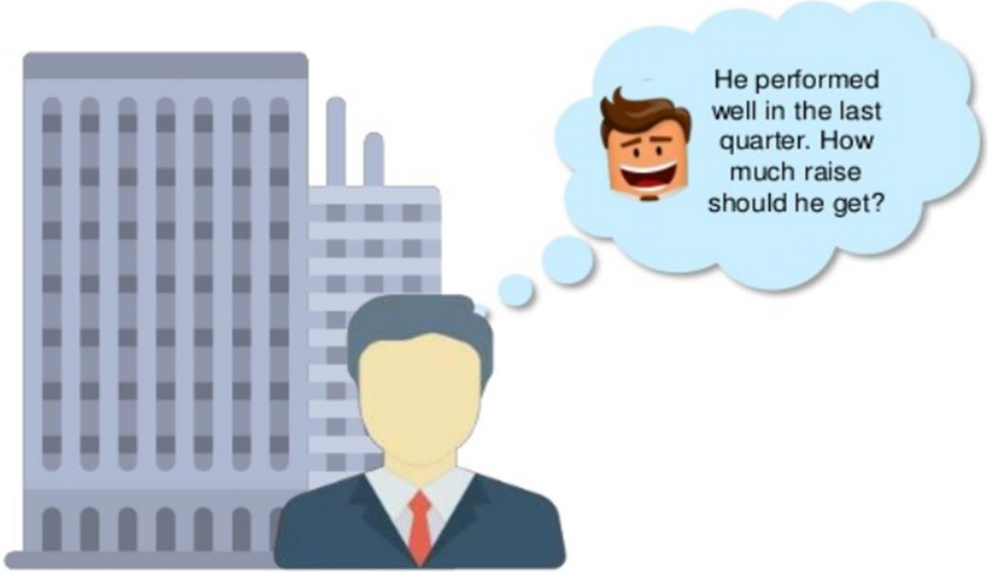


Model makes predictions

Logistic Regression

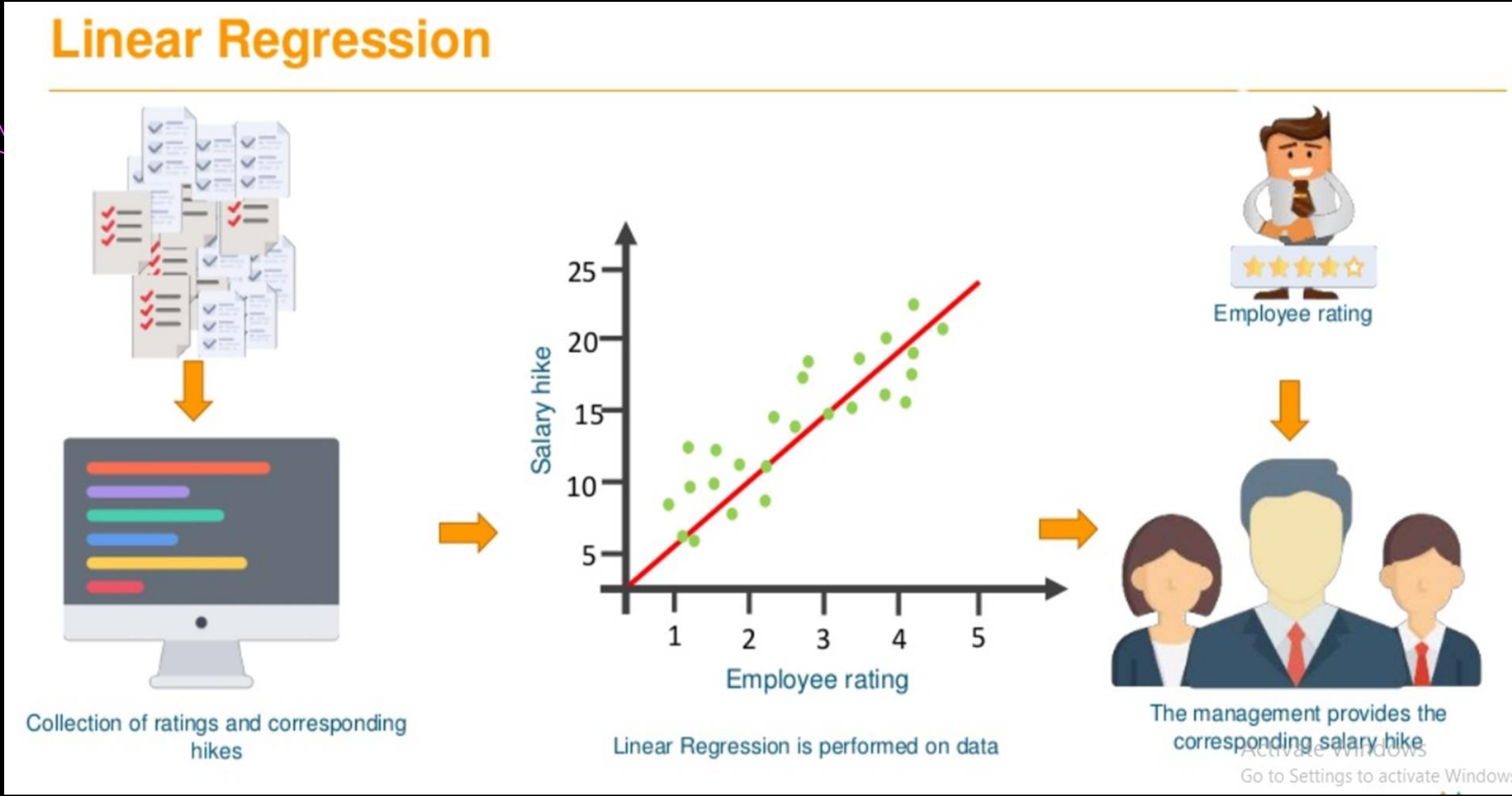
Logistic Regression

Linear Regression



It is a statistical method that helps find the relationship between an independent and dependent variable, both of which are continuous

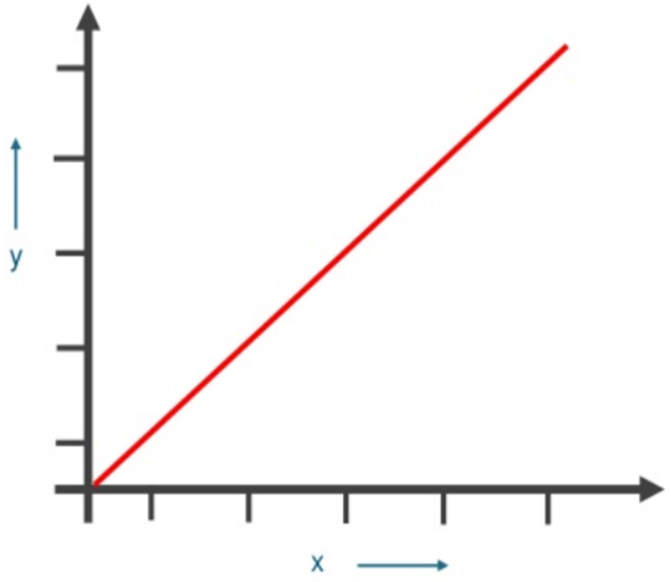
Logistic Regression



Logistic Regression

Linear and Logistic Regression

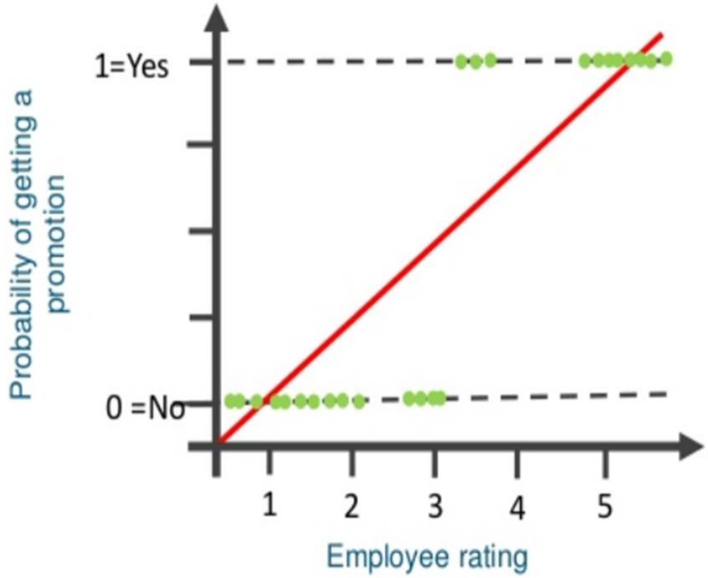
Here's the graph of how linear regression would be, for a given scenario



Logistic Regression

Linear and Logistic Regression

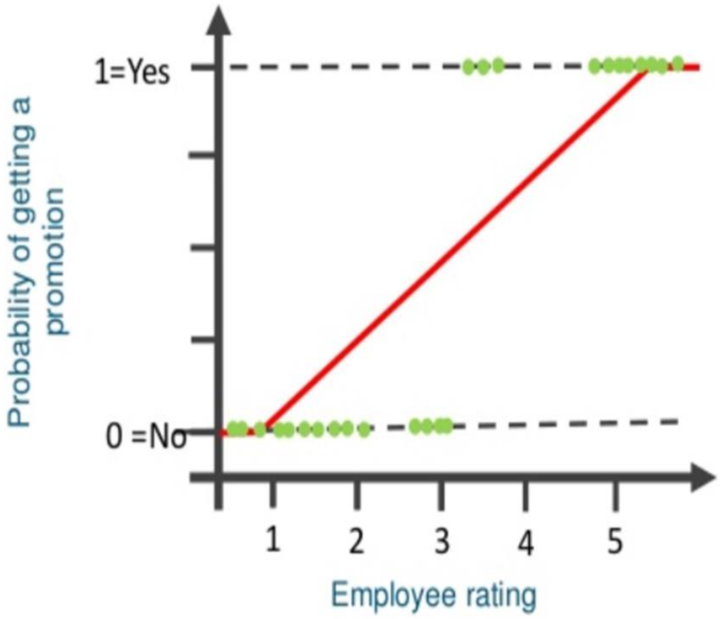
What if you wanted to know whether the employee would get a promotion or not based on their rating



Logistic Regression

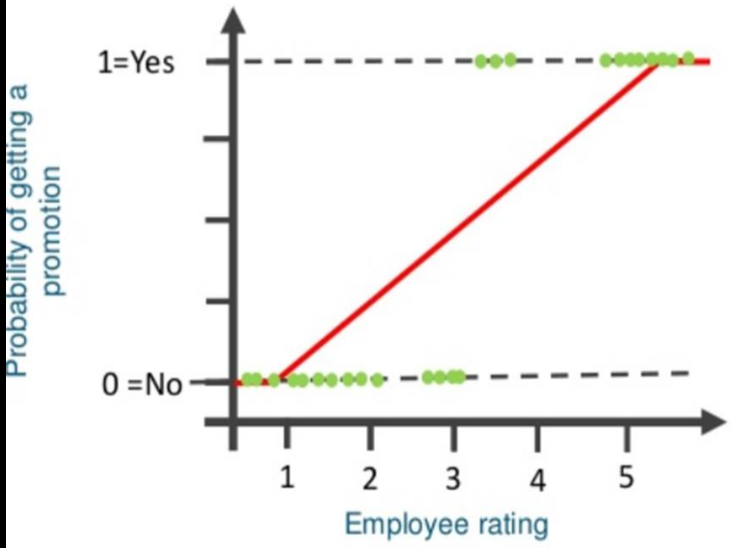
Linear and Logistic Regression

This graph would not be able to make such a prediction. So we clip the line at 0 and 1.



Logistic Regression

Linear and Logistic Regression



So, how did this…

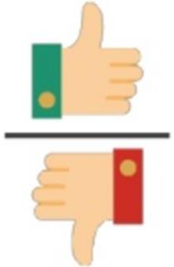
…become this?

Logistic Regression

The Math behind Logistic Regression

To understand Logistic Regression, let's talk about the odds of success

Odds (θ) =



$$\frac{\text{Probability of an event happening}}{\text{Probability of an event not happening}} \quad or \quad \theta = \frac{p}{1-p}$$

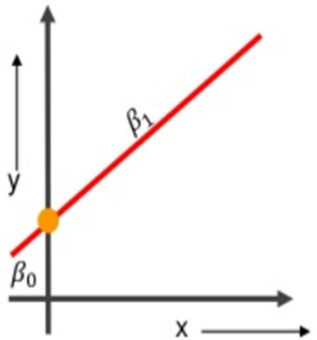
The values of odds range from 0 to ∞

The values of probability change from 0 to 1

Logistic Regression

The Math behind Logistic Regression

Take the equation of the straight line



Here, β_0 is the y-intercept
β_1 is the slope of the line
x is the value of the x co-ordinate
y is the value of the prediction

The equation would be: $y = \beta_0 + \beta_1 x$

Logistic Regression

The Math behind Logistic Regression

Now, we predict the odds of success

$$\log\left(\frac{p(x)}{1-P(x)}\right) = \beta_0 + \beta_1 x$$

Exponentiating both sides:

$$e^{ln\left(\frac{p(x)}{1-p(x)}\right)} = e^{\beta_0+\beta_1 x}$$

$$\left(\frac{p(x)}{1-p(x)}\right) = e^{\beta_0+\beta_1 x}$$

Let Y = $e^{\beta_0+\beta_1 x}$

Then $\frac{p(x)}{1-p(x)}$ = Y

$$p(x) = Y\big(1 - p(x)\big)$$

$$p(x) = Y - Y\big(p(x)\big)$$

$$p(x) + Y\big(p(x)\big) = Y$$

$$p(x)(1 + Y) = Y$$

$$p(x) = \frac{Y}{1+Y}$$

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

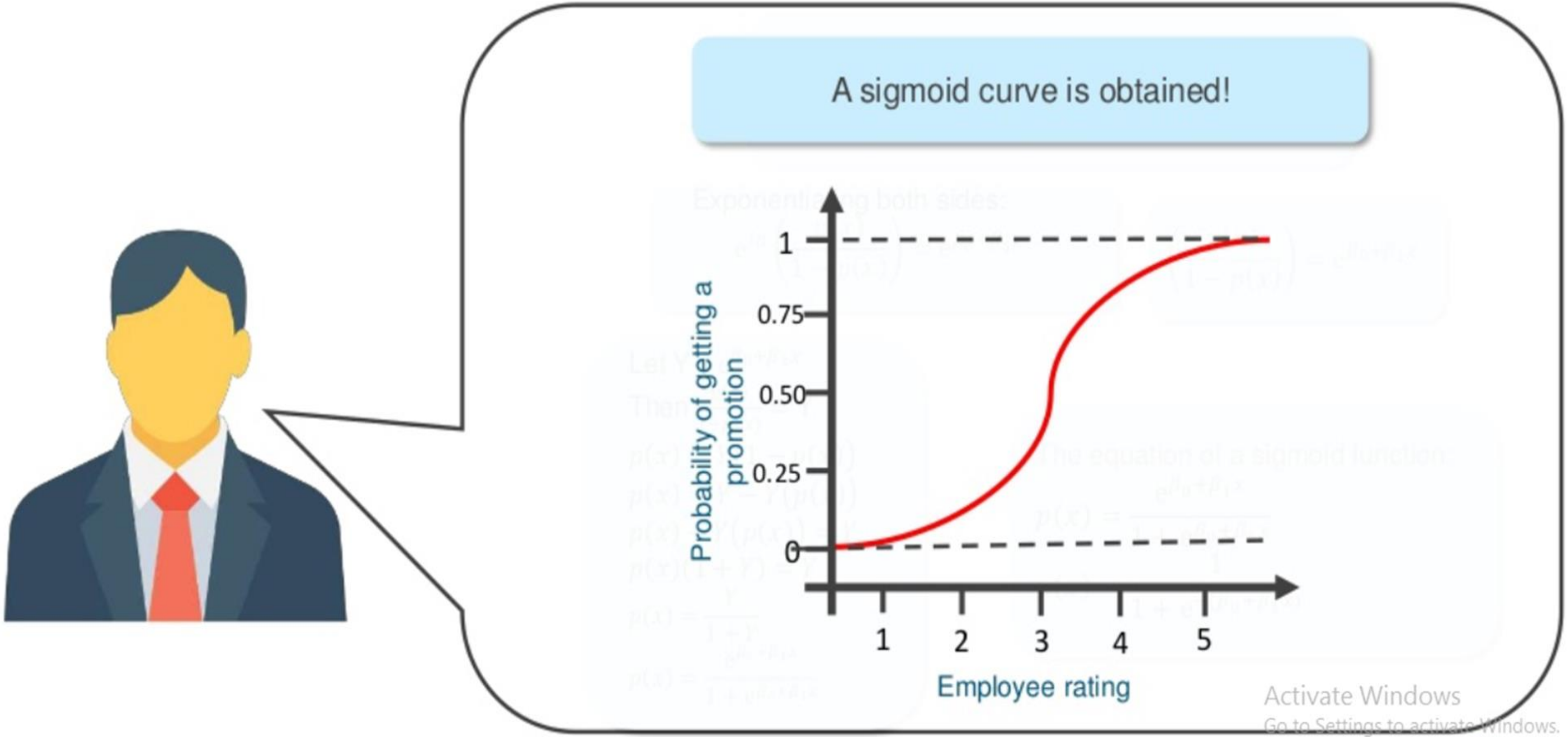
The equation of a sigmoid function:

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$\boxed{p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}$$

Logistic Regression

The Math behind Logistic Regression



Logistic Regression

Logistic Regression

How is Linear and Logistic Regression different?

Linear Regression

- Used to solve Regression Problems
- The response variables are continuous in nature
- It helps estimate the dependent variable when there is a change in the independent variable.
- Is a straight line.

Logistic Regression

- Used to solve Classification Problems
- The response variable is categorical in nature
- It helps calculate the possibility of a particular event taking place.

Logistic Regression

How is Linear and Logistic Regression different?

Linear Regression	Logistic Regression
• Used to solve Regression Problems • The response variables are continuous in nature • It helps estimate the dependent variable when there is a change in the independent variable. • Is a straight line.	• Used to solve Classification Problems • The response variable is categorical in nature • It helps calculate the possibility of a particular event taking place. • An S-curve. (S = Sigmoid)

Logistic Regression

Logistic Regression Applications

Weather Prediction

Helps determine the kind of weather that can be expected

Logistic Regression

Logistic Regression Applications

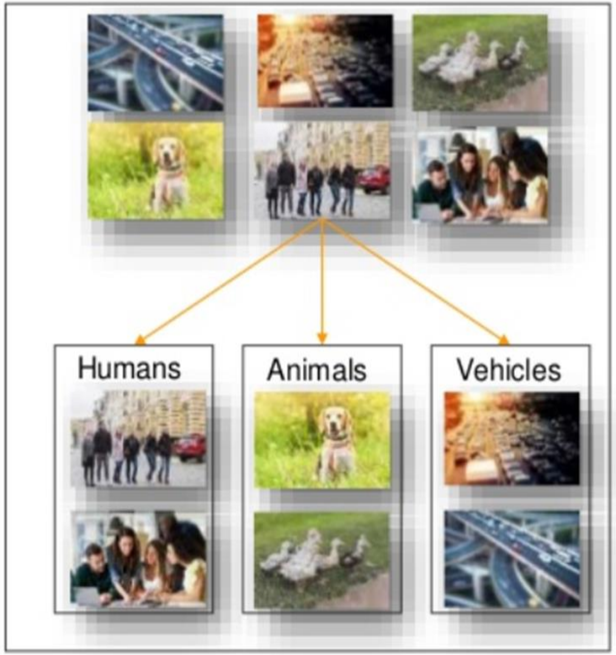
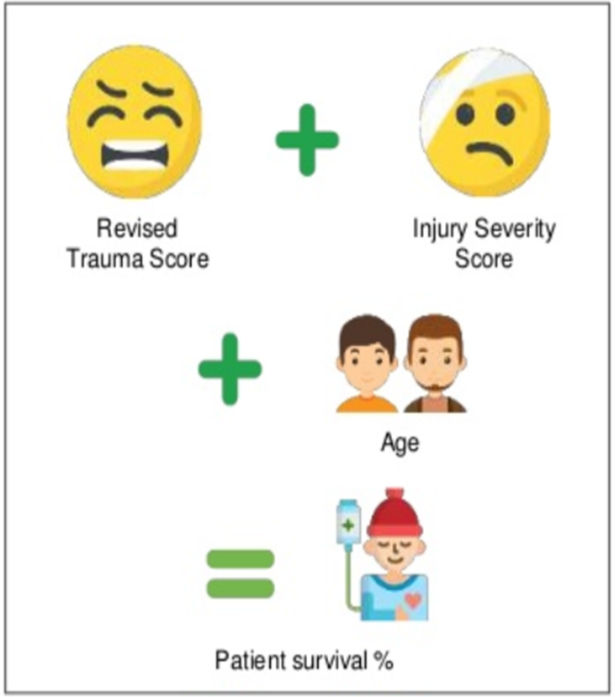


Image Categorization

Identifies the different components that are present in the image, and helps categorize them

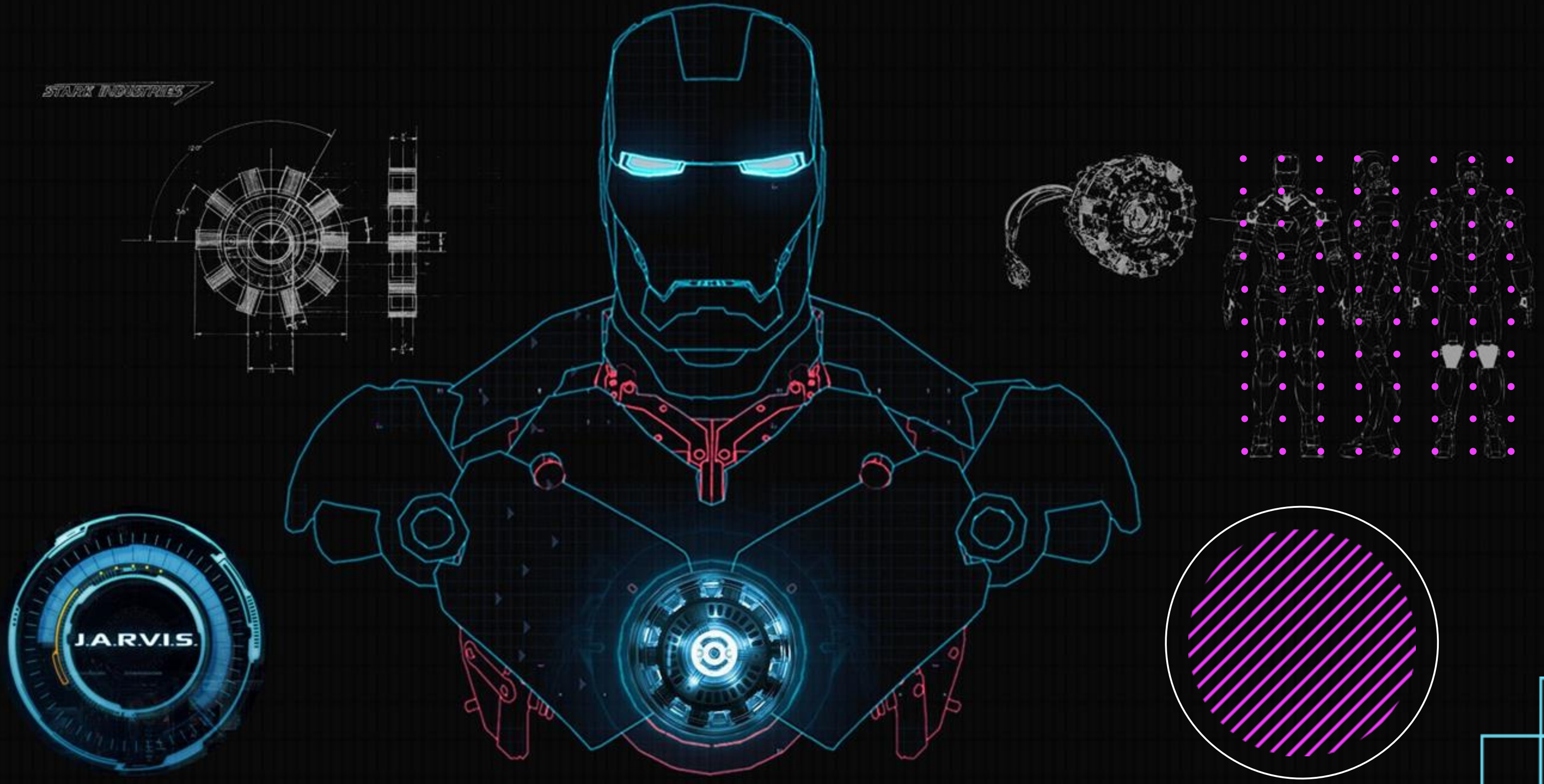
Logistic Regression

Logistic Regression Applications



Healthcare (TRISS)

Determines the possibility of patient survival, taking age, ISS and RTS into consideration



Thank You