10 February 2024 20:04

Who am I ?

01 October 2023 08:18

AKASH PUSHKAR CHARAN *aka* APC

Quick witted | Tech-Savvy | Observer

- **IIT Kanpur Alumnus**
- **Working as Lead Data Scientist with Accenture Strategy & Consulting**

8+ *of Industry experience of delivering* **multiple data science projects** *across industries*

15+ *years of experience training & mentoring*

Taught **30000+** *GATE Students*

5000+ *career transitions into Data Science roles*

"It's not who I am underneath but what I do that defines me"

IIT-JEE MATHS since 2008

GATE - ECE -
Data Science /ML -
2024 -

https://www.linkedin.com/in/akash-pushkar-04642925/

Setting the ground rules and expectations

10 February 2024 20:14

Ground rules:

① Sheer focus, attention
↳ Focus
DND

② **Revision** is the key to machine learning
↳ Always revise all the notes (One Note)
- notebooks
(gtg: good to go) ↳ Jupyter Notebook
• ipynb.

Doubt:
↳ Slot #1: After the break.
↳ Slot #2: After the session ends.

UNMUTE

③ Industrial Use cases.
↳ good picture about what happens in a real projects.

Any topic ✗ : Linear Regression.

↳ Intro about LR
↳ Intuition
↳ Mathematics → good idea → In-depth.
(Pre-requisites)
a) Calculus optimization
b) Linear Algebra
c) Probability & statistics
behind algo
↳ Python Hands-on.

COURSE FLOW

10 February 2024 20:34

1. Introduction to Machine learning
2. Linear Regression
3. Logistic Regression
4. Decision Trees
5. Random Forest

supervised learning

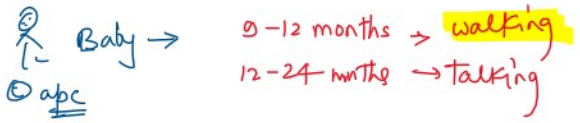
6. K-Means clustering
7. Hierarchical Clustering

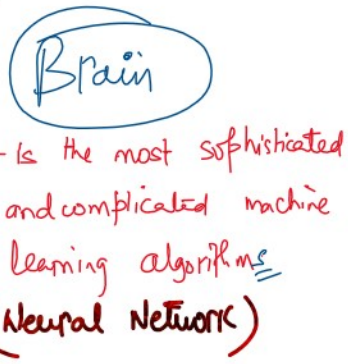
Unsupervised learning

Overview - Machine Learning

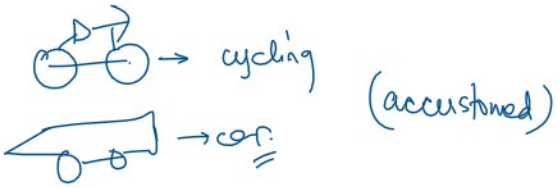
10 February 2024 20:37

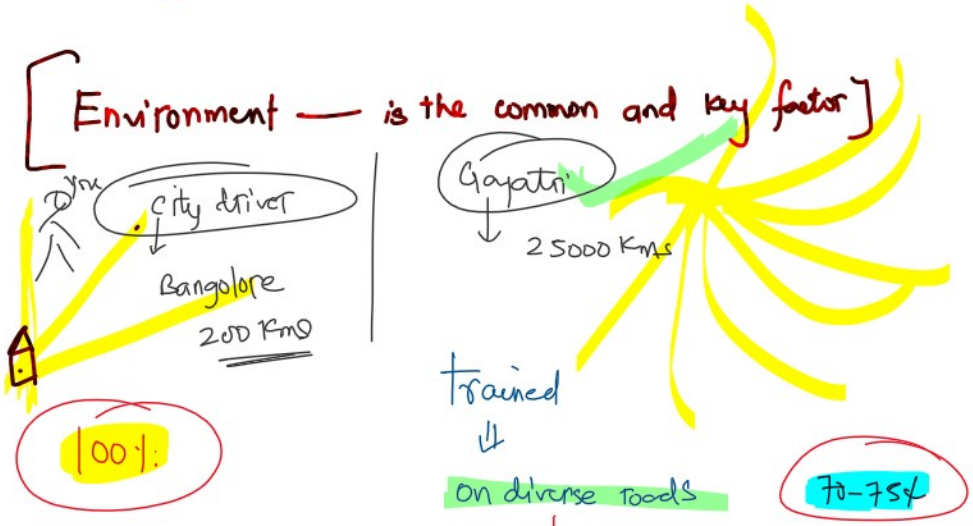
Q: How would you explain machine learning to 5 year old kid??





Driving:





(ML + optimization + Simultation)

Google Newsstand

India Today
COP28 Dubai Big Failure: Why Countries Can't Agree On Phasing Out Fossil Fuels
10 hours ago

(Clustering + NLP

Business Standard

Making history, COP28 nations agree to 'transition away' from fossil fuels

9 hours ago • Shreva Jai

India Today

COP28 Dubai Big Failure: Why Countries Can't Agree On Phasing Out Fossil Fuels

10 hours ago

European Commission

Daily News 13 / 12 / 2023

15 hours ago

FE The Financial Express

A balanced consensus: Despite scepticism, UAE has delivered a balanced package on climate action at COP28

3 hours ago • Opinion • Chandra Bhushan

Full coverage

(+ NLP)

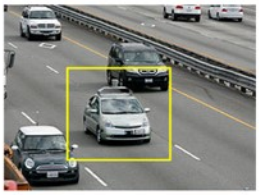
{ Alexa | Google Home | Siri # etc }

↳ (NLP + NLG)

(AI based Camera) ✓✓

Tesla / BYD

Autonomous Cars

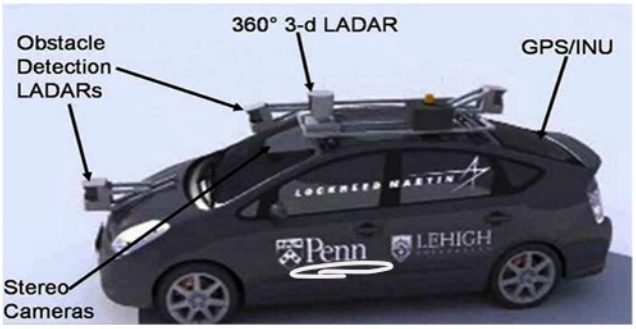


- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Penn's Autonomous Car → (Ben Franklin Racing Team)

12

Autonomous Car Sensors



which type of model is created in ML for autonomous cars

-- COMPUTER VISION

What is Machine Learning?

"Learning is any process by which a system improves performance from experience."

- Herbert Simon

Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E.

A well-defined learning task is given by <P, T, E>.

Spam Mail Detection

P: How accurately Categorize email messages as spam or ham accurately.
— % of email correctly classified, ↓ important or not a spam.

T: Spam vs Ham Mail Tagging

E: Database of emails, + User feedback.

(P)

100 mails

80 mails (Ham)

20 mails (Spam)

(15 mails were actually spam)

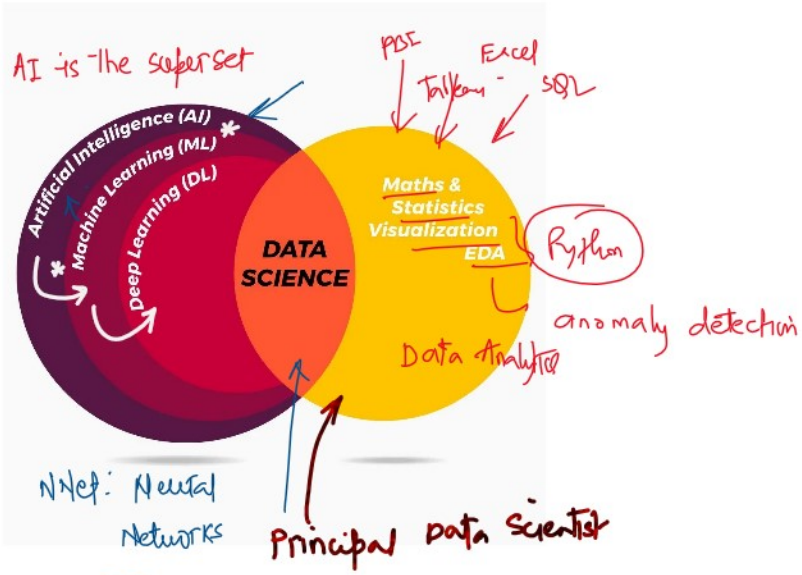
accuracy = $\frac{15}{20} \times 100 = 75\%$.

Reading Assignment :

HW ① How does GMAIL classify mail as spam vs ham ?

AI | ML | DS → Data Science.

DL : Deep learning



CNN

Key areas of Machine learning

(80-90%)

30% **Prediction** —Machine learning can also be used in the prediction systems. Considering the loan example, to compute the probability of a default, the system will need to classify the available data in groups.

10% **Image recognition** —Machine learning can be used for face detection in an image as well. There is a separate category for each person in a database of several people.

Speech Recognition —It is the translation of spoken words into the text. It is used in voice searches and more. Voice user interfaces include voice dialing, call routing, and appliance control. It can also be used a simple data entry and the preparation of structured documents.

Medical diagnoses —ML is trained to recognize cancerous tissues.

Financial industry and trading— companies use ML in fraud investigations and credit checks.

[Weekly sales (Time series Forecasting)], [Weather forecasting]

[Predictive Analytics]

[Fraud detection] → (HP Internal Audit)

Preventing it → [Fraud Prevention]

Indore → attending a function ??

Gift

(Merchant)

Financial industry and trading— companies use ML in fraud investigations and credit checks.

(A) How much product 'X', walmart store will be selling in the next week ?
(100 or 80 units or 0 units)

Let us say product 'X' is a new product !

(B) Whether product 'X' will sell or not ??

Yes No (Binary)

Supervised learning

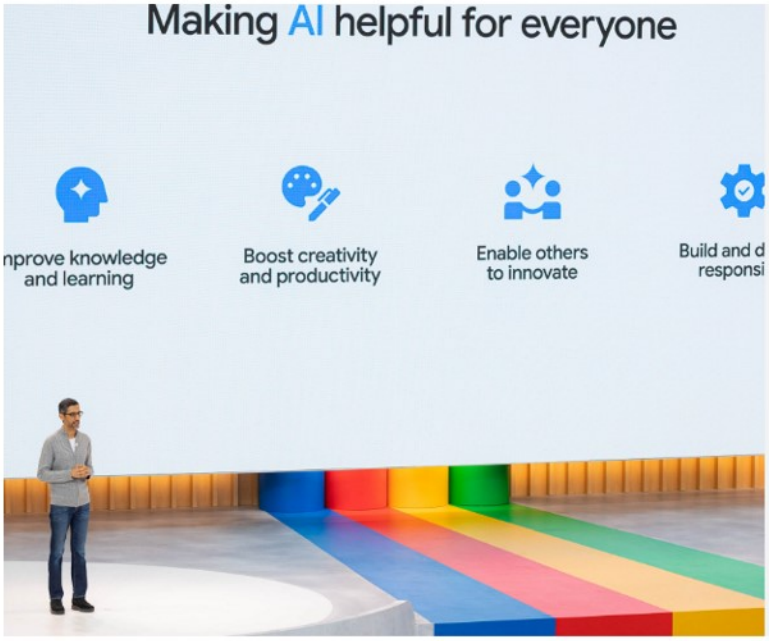
Regression → A

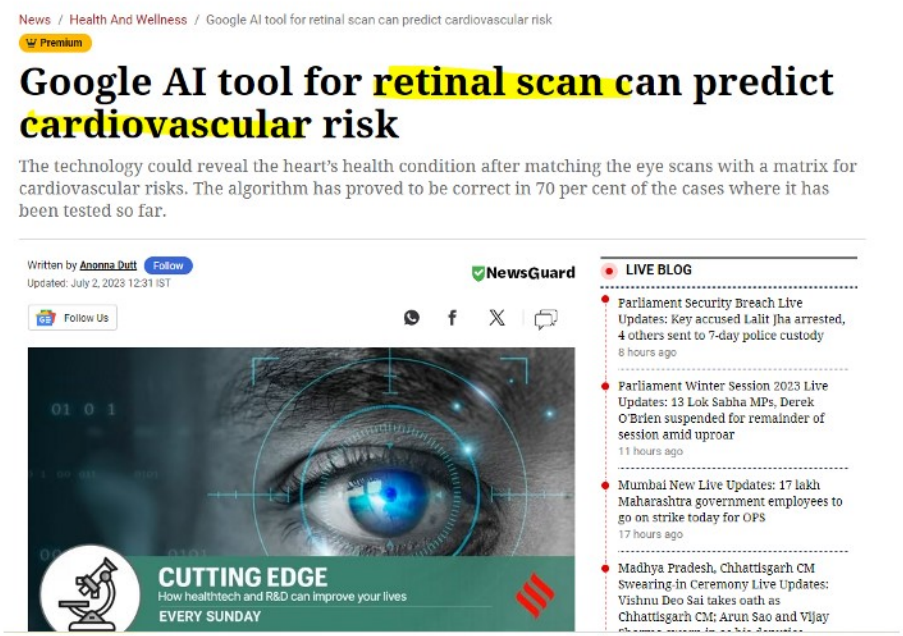
Classification → B

How much ⇒ Regression

Whether Yes | No ⇒ Classification

"supervised learning" → 90% of good data scientists





News / Health And Wellness / Google AI tool for retinal scan can predict cardiovascular risk

Premium

Google AI tool for retinal scan can predict cardiovascular risk

The technology could reveal the heart's health condition after matching the eye scans with a matrix for cardiovascular risks. The algorithm has proved to be correct in 70 per cent of the cases where it has been tested so far.

Written by Anonna Dutt Follow
Updated: July 2, 2023 12:31 IST

Follow Us

NewsGuard

LIVE BLOG

- Parliament Security Breach Live Updates: Key accused Lalit Jha arrested, 4 others sent to 7-day police custody
  8 hours ago
- Parliament Winter Session 2023 Live Updates: 13 Lok Sabha MPs, Derek O'Brien suspended for remainder of session amid uproar
  11 hours ago
- Mumbai New Live Updates: 17 lakh Maharashtra government employees to go on strike today for OPS
  17 hours ago
- Madhya Pradesh, Chhattisgarh CM Swearing-in Ceremony Live Updates: Vishnu Deo Sai takes oath as Chhattisgarh CM; Arun Sao and Vijay

Machine Learning :

There are 4 major techniques:

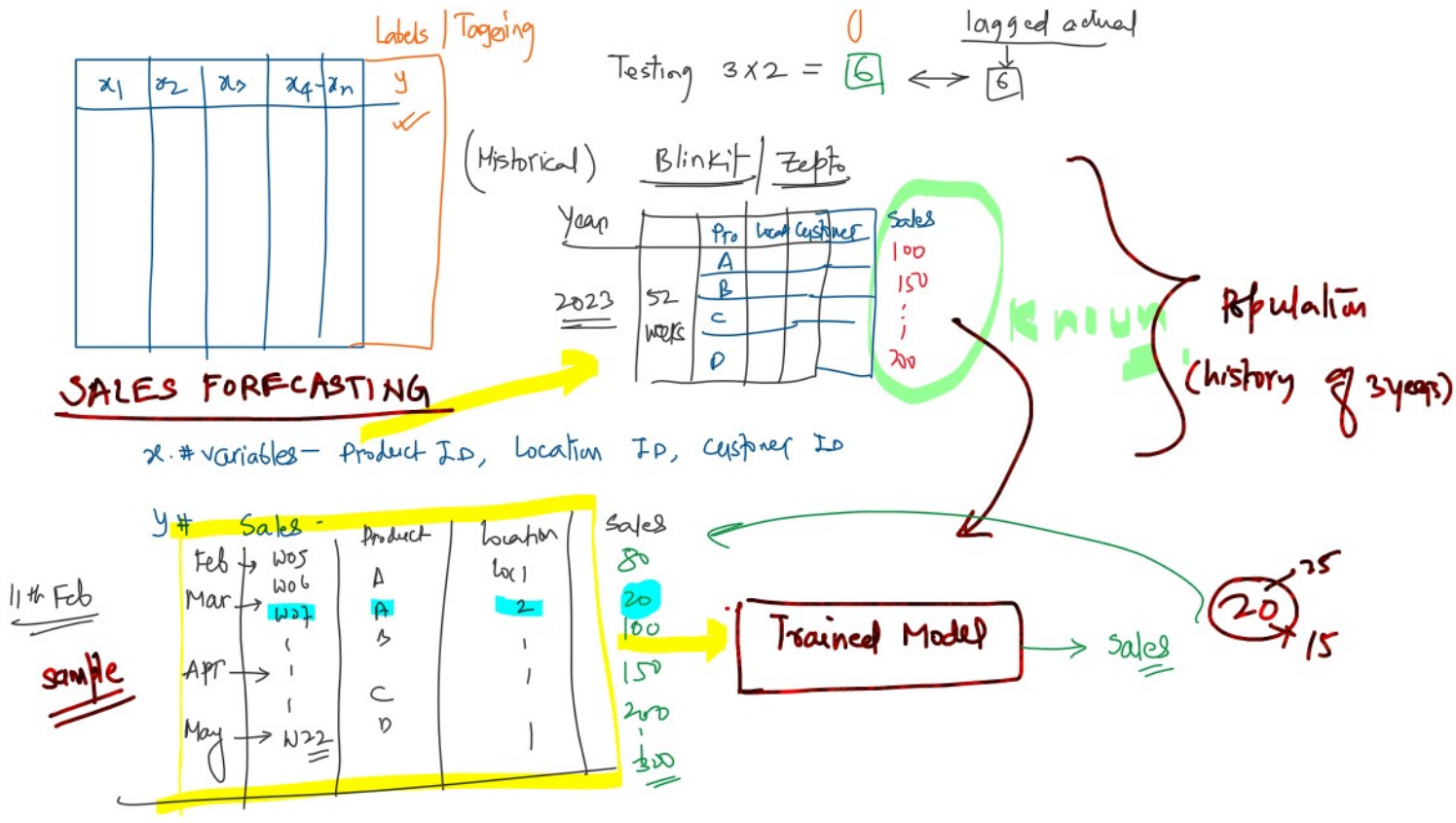
1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Semi - supervised.

Training $2 \times 3 = 6$ (x_1, x_2, y)

Testing $3 \times 2 =$ 6 $\leftrightarrow$ 6 (Tagged actual)

| x_1 | x_2 | x_3 | x_4 | x_n | y |

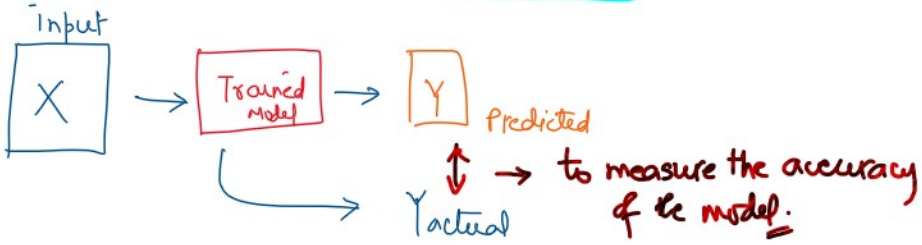
Labels | Tagging



Overview of Supervised Learning Algorithm

In Supervised learning, an AI system is presented with data which is labelled, which means that each data tagged with the correct label.

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.



Types of Supervised Learning

depends on Y

- Regression : Sales Forecasting is a regression problem. | Y is continuous ⇒ Regression
  - Linear Regression
- Classification : Spam mail detection is a classification problem | Y is categorical ⇒ Classification
  - Logistic Regression

Regression: derived from the latin word "regredi" which means go back or to return.

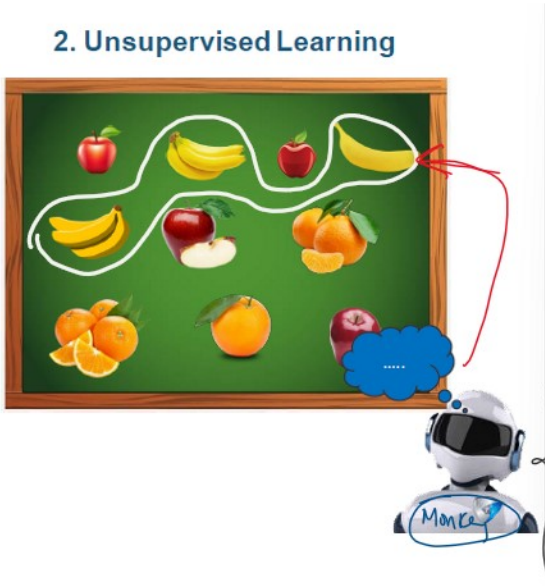
means go back or to return.

→ Sir Francis Galton – 19th century coined the term regression

Regression: A regression problem is when the output variable is a real value, such as "dollars" or "weight". (continuous)

Classification: A classification problem is when the output variable is a category, such as "red" or "blue" or "disease" and "no disease". (discrete | categorical)

Overview of Unsupervised Learning Algorithm

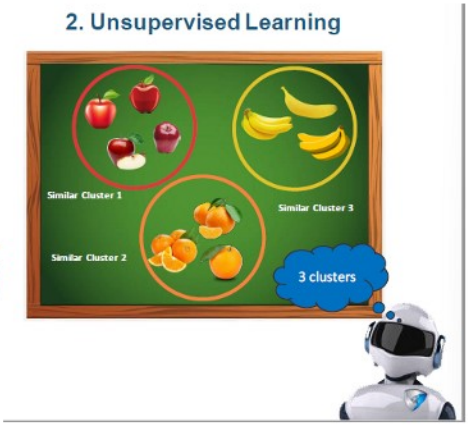


Banana

Neilsen

Just to find the similarity

shape — x_1
size — x_2
color — x_3
texture — x_4
features



Customer Segmentation

– collecting demographic data

CC#1 < 4 lpa

CC#2 4 < 10 lpa

credit card

↳ credit limit ↑ 30%

credit score

↓

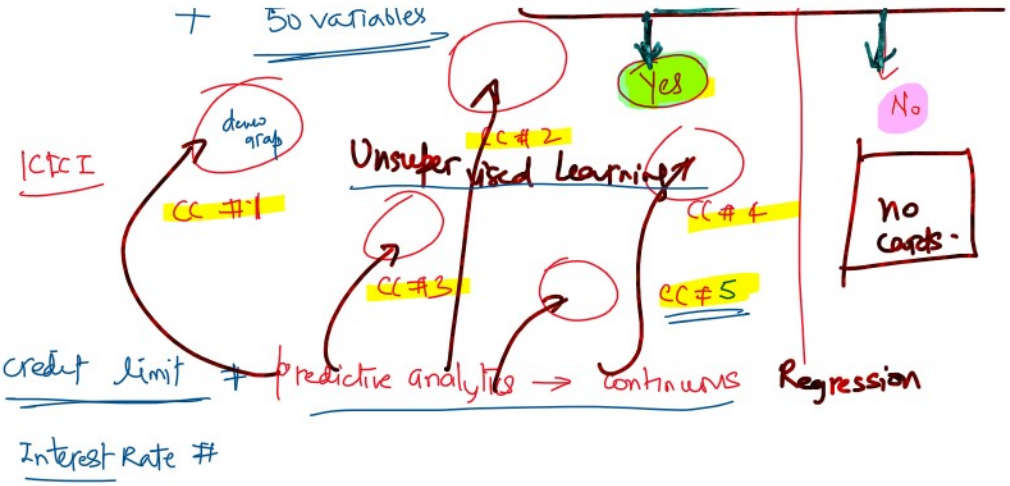
CIBIL

→ Unsupervised learning

Age, Gender, Income, city, any previous loan, payments defaulted

+ 50 variables

classification



↳ Increase credit limit

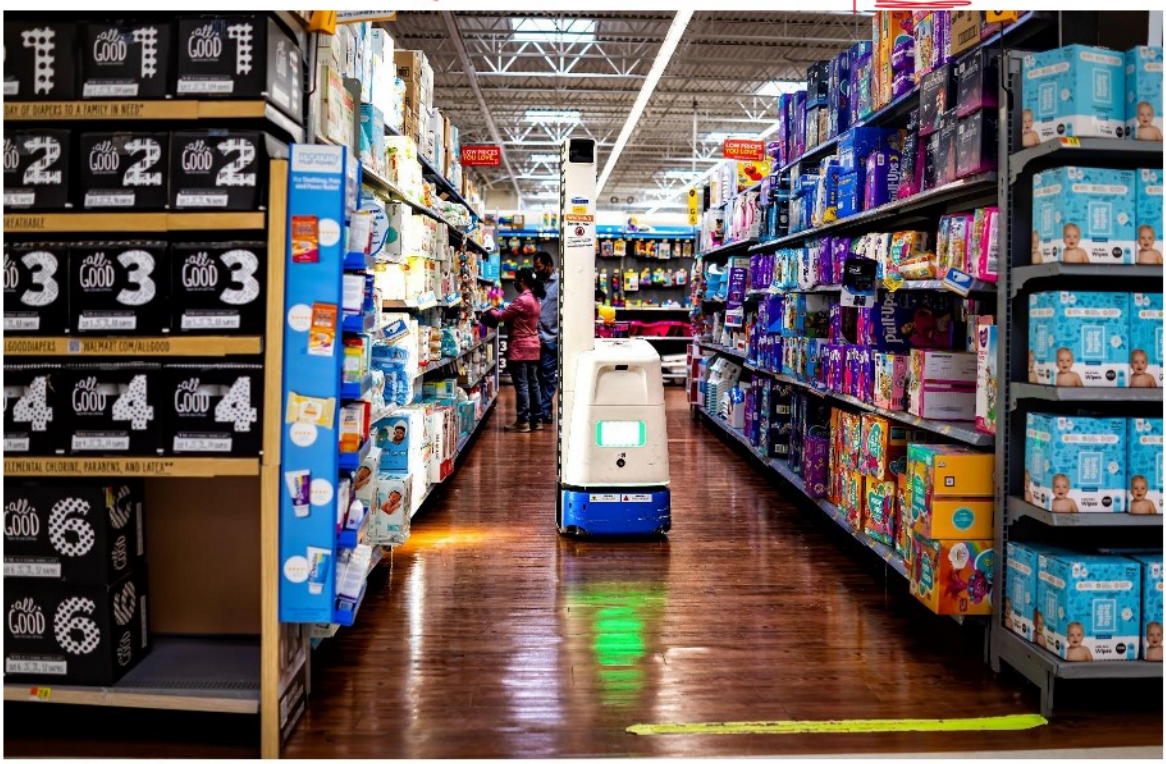
↳ Upgrade the existing card

↳ Waive off annual charges | late charges

Types of Unsupervised learning

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior. ~" K-Means clustering"
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

(increased sales) WalMart | DMart

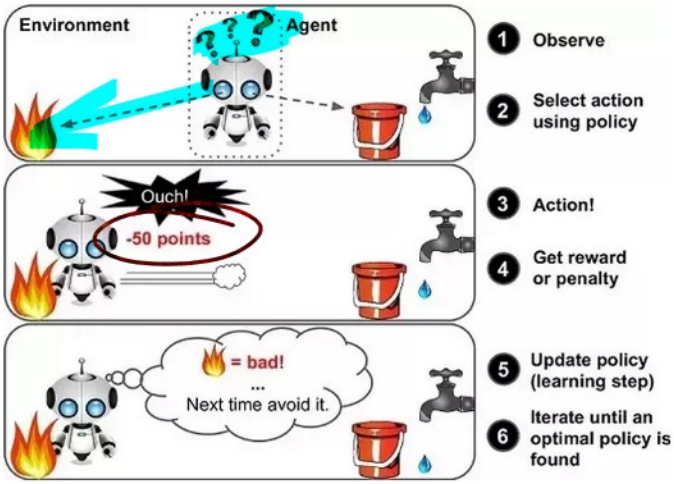


https://bigdatabigworld.wordpress.com/2014/11/25/beer-and-nappies/

https://canworksmart.com/diapers-beer-retail-predictive-analytics/

https://tdwi.org/articles/2016/11/15/beer-and-diapers-impossible-correlation.aspx

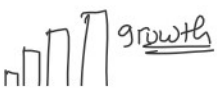
Overview of Reinforcement Learning



- A reinforcement learning algorithm, or agent, learns by interacting with its environment.
- The agent receives rewards by performing correctly and penalties for performing incorrectly.
- The agent learns without intervention from a human by maximizing its reward and minimizing its penalty. It is a type of dynamic programming that trains algorithms using a system of reward and punishment.

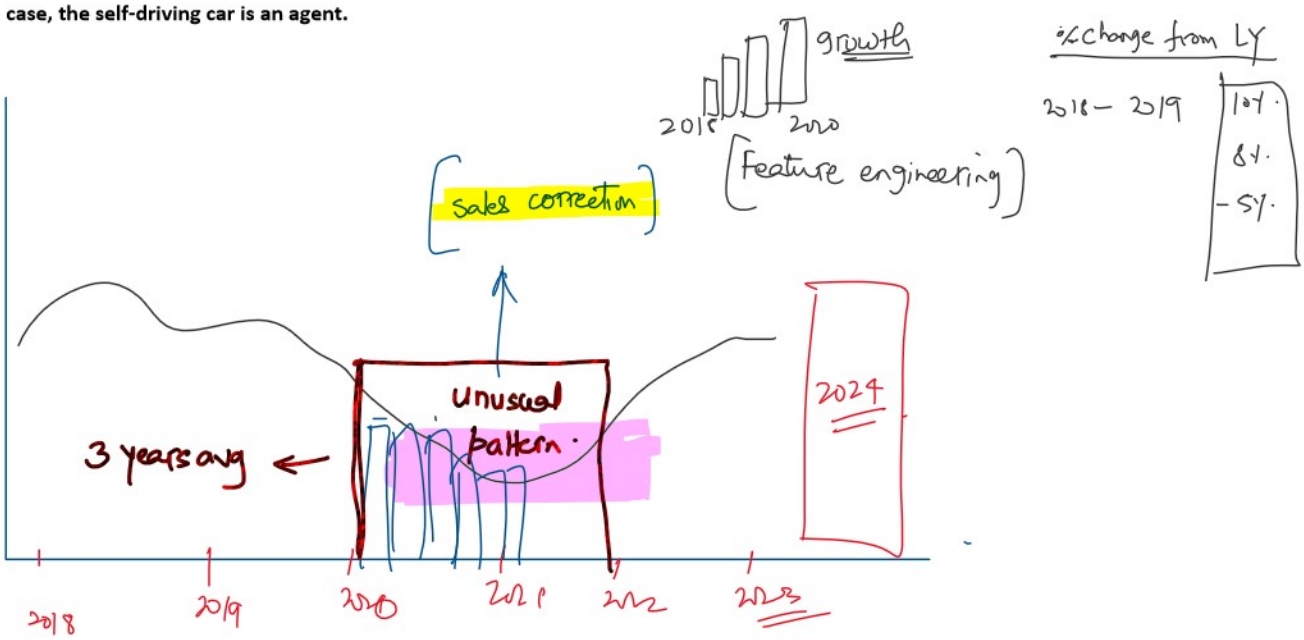
– Reinforcement learning is used for self-driving cars.

Reinforcement learning (RL) is a type of machine learning where an agent learns by exploring and interacting with the environment. In this case, the self-driving car is an agent.



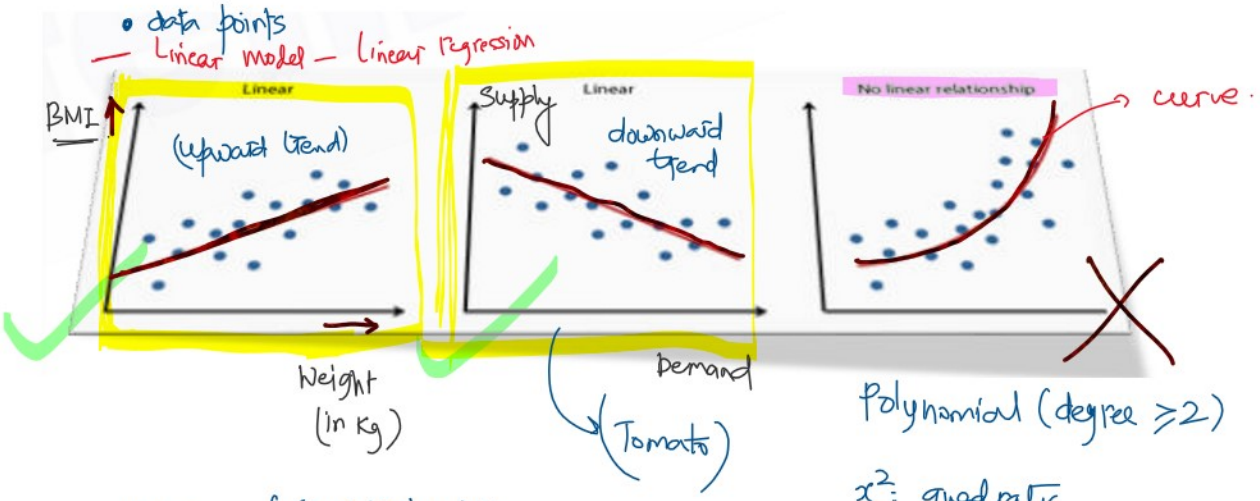
% change from LY

case, the self-driving car is an agent.

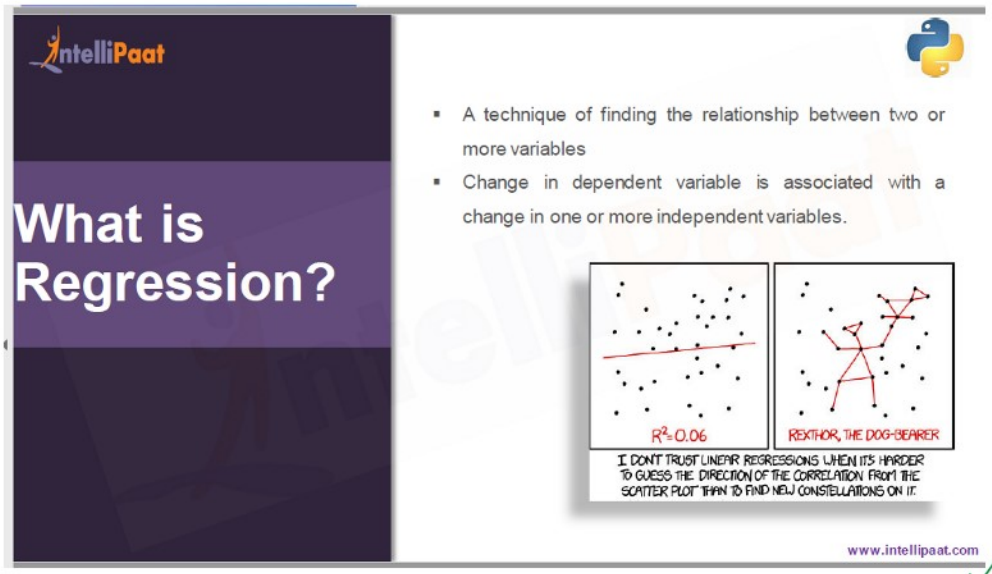


Linear Regression

11 February 2024 22:15

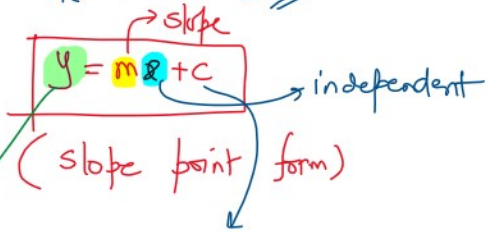


$BMI = f(weight, height)$



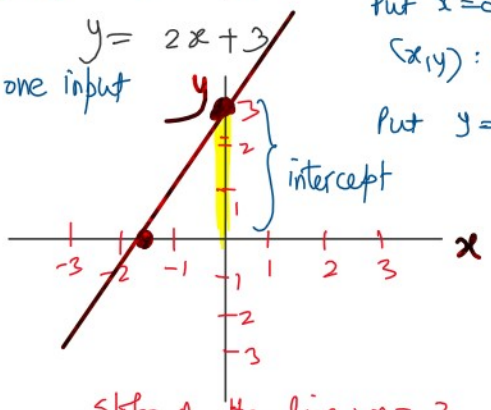
$y = f(x_1, x_2, x_3 \cdots x_n)$

- a functional mapping
- to find the relationship between y and two or more x variables.



[Simple Linear Regression]

it has only one input



Put $x=0$, $y = 2\times0+3 = 3$

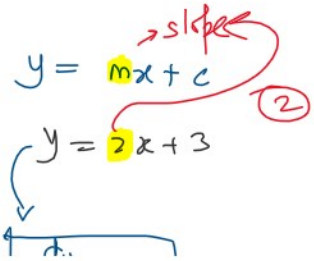
$(x, y) : (0, 3)$

Put $y=0$, $x = -\frac{3}{2} = -1.5$

$(x, y) : (-\frac{3}{2}, 0)$

slope of the line : $m = 2$

intercept of the line : $c = 3$



Slope of the line: $m = 2$

intercept of the line: $c = 3$

(Multiple Linear Regression)

$$y = 3 + 2x_1 + 5x_2 + 3.7x_3 - 0.8x_4$$

$$\frac{dy}{dx} = 2$$

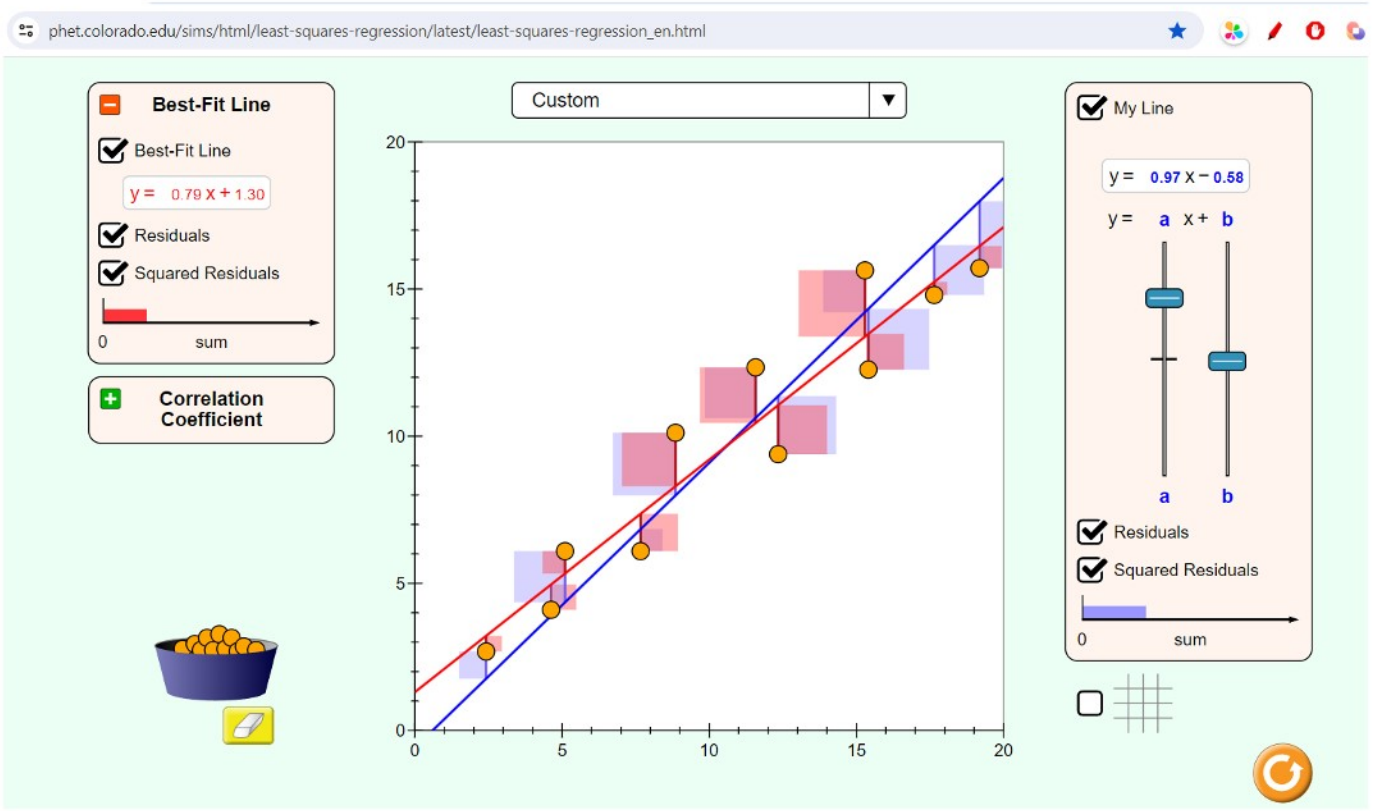
First order derivative is slope which is 2.

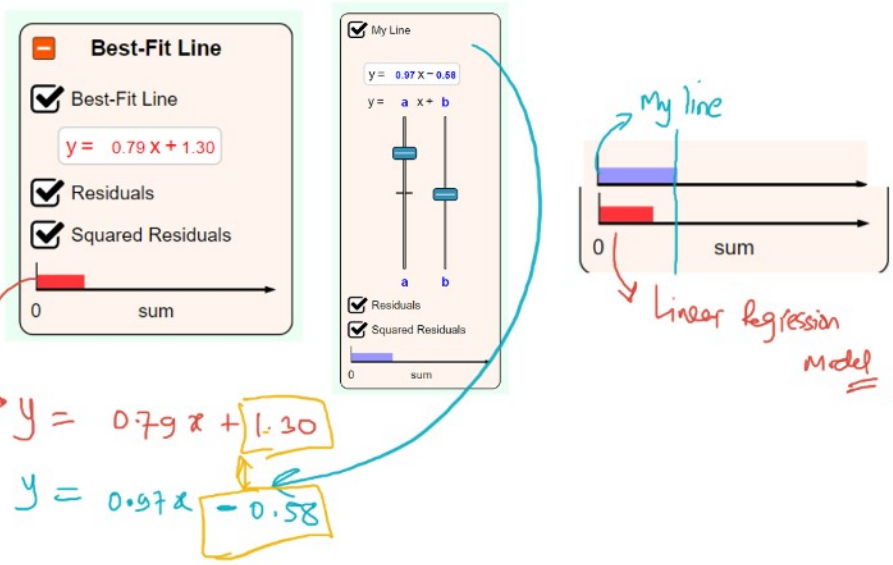
Notations

x : input variable | features | independent | predictor

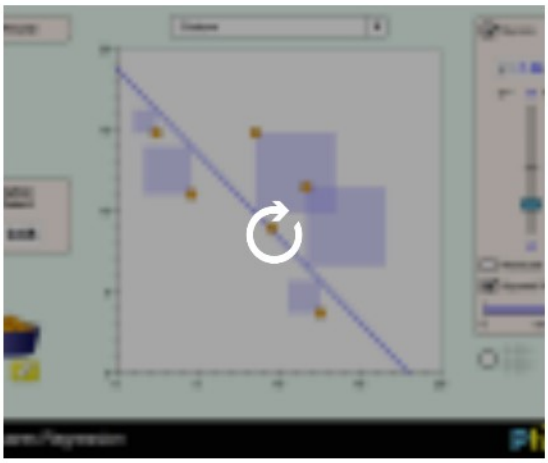
y : output variable | target | dependent | response

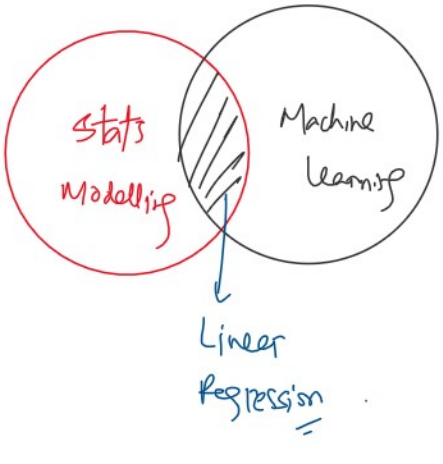
Intuition behind Linear Regression:

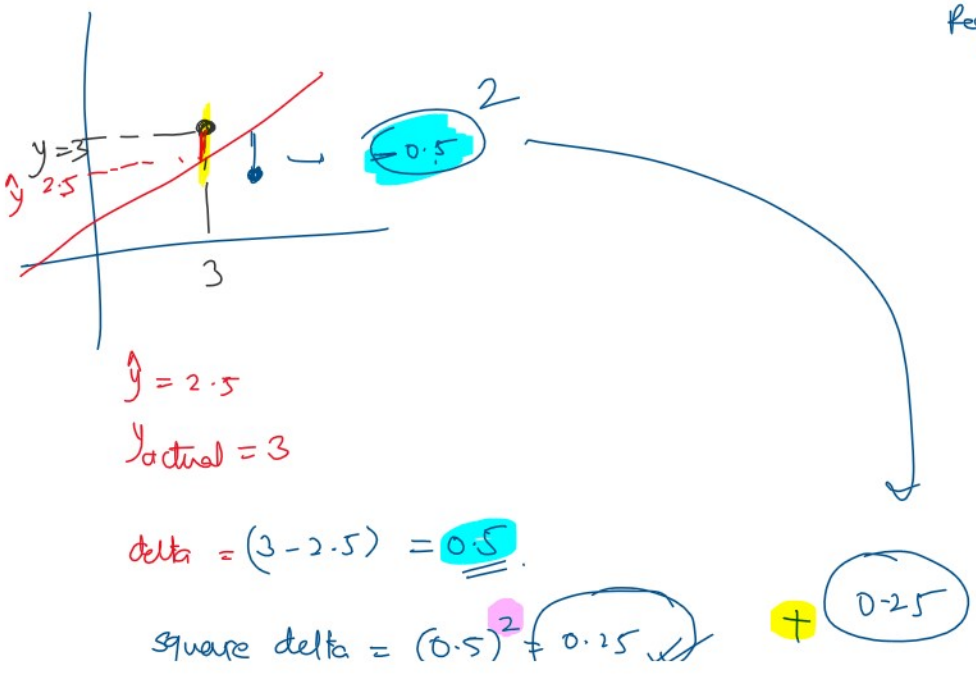




[Least-Squares Regression]







square delta = $(0.5)^2 = 0.25$ + 0.25

Sum of square residuals

$y = mx + c$ (linear)

$m = -2$

$2x + y = 18$

18

$y = -2x + 18$