WELCOME :)

STARTS AT 9:05

INTRO TO CLASSICAL ML

SORRY FOR SMALL SCREEN :/

- Already Learnt - **Summary**
- To Learn - **Overview**

Libraries

DSA

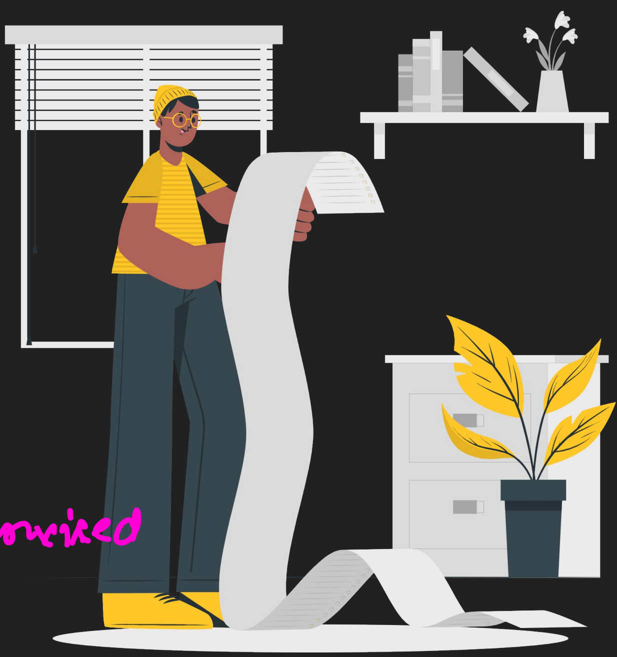
Linear Algebra

1. What is ML ? - Tom Mitchell
2. ML v/s SDE Coding
3. ML Tasks

   Regression

   Classification
4. Types of Learning

Historical Data : Supervised/Unsupervised



Let's Recall . . .

Let's briefly discuss what all we learnt in previous modules

1. DSML Libraries
2. Probability and Statistics
3. Coordinate Geometry & Linear Algebra
4. Calculus and Optimization

DSML Libraries

mp, pd, plt, sns

Which plot to make?

→ # variables

→ Types of v - cont, cat (Num)

1. Data loading, processing and manipulation - **NumPy, Pandas**
2. Plotting - **matplotlib, pyplot, Seaborn**

↓

pyplot

Let's revise some key thumb rules we can use figure out which plot to use

DSML Libraries - Cutting the clutters.

UNIVARIATE DATA VIZ.

1V

Numerical - histogram, KDE

→ Distribution of data

→ Statistics of the data ???? Mean, Med, Min Max, 75% 25th

Categorial - bar-plot, count-plot

→ Frequency Dist

→ Proportion.

BIVARIATE

2V

Numerical - Numerical - ~~histogram, KDE~~

→ Association b/w two variable.

→ Quantifying the associations - Correlation

Categorical - Categorical - histogram, KDE

→ Distribution of categories within categories

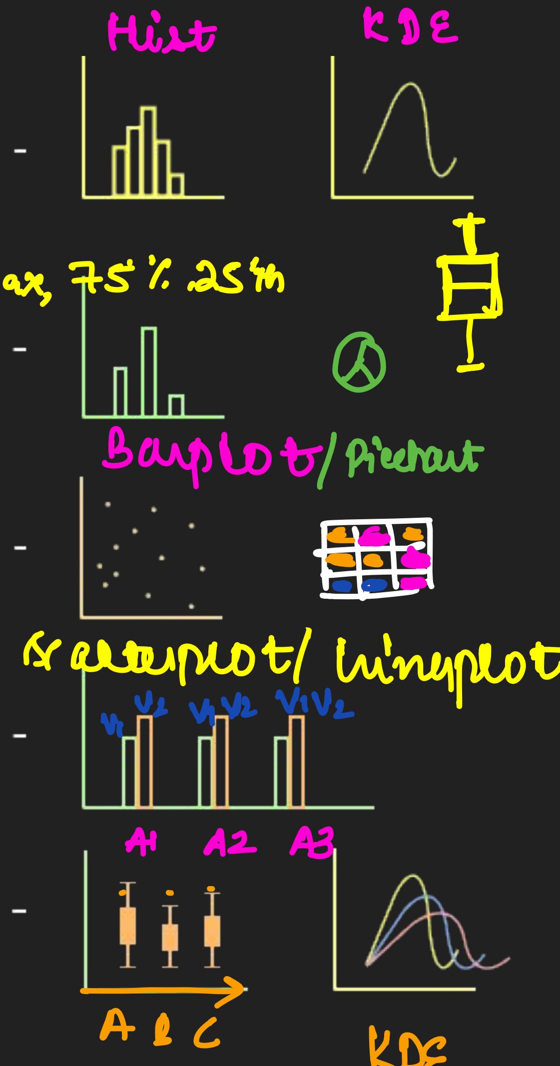
→ stacked / Dodged Bar plot

Numerical - Categorial - bar-plot, count-plot

One Number - Barplot

Multiple Numbers - Boxplot

Distribution - KDE with hue



Plotting Examples :

Task →

Patient(X) - Is diabetic ??

w h

features / Variables

continuous continuous.

3rd variable

Categorical.

Which plot can we use here ? Scatter plot

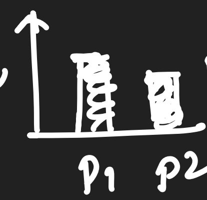
Addition : Mark every scatter point as D/ND

Is there a way we can represent the third feature is_diabetic on this plot?

color data pts based on category

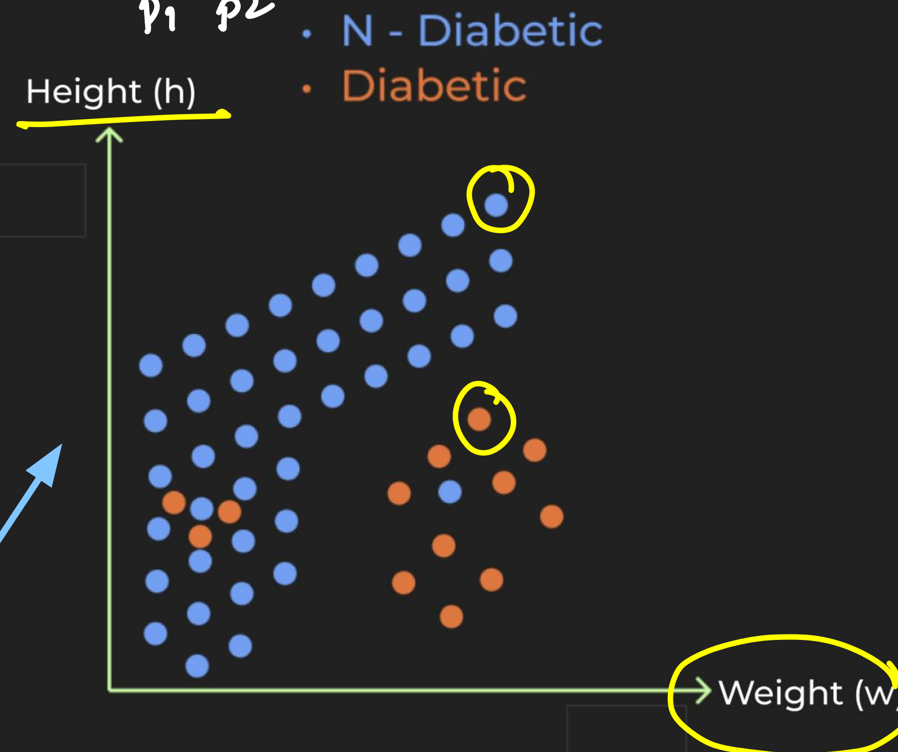
· Products - 1 and 2 - categorical

- continuous



sales S1

sales S2



Technically, we have plotted three variables on this 2D plot

- numerical - numerical - categorical

3V

Multivariate Data Visualisation

Probability & Statistics : Diabetic / N-Diabetic (W) (H)

Classification - D/ND

Output → P(diabetic)

1. Remember we learnt how can we calculate things like P(diabetic)

W, H → D / ND

P(diabetic)

Probabilistic Angle

2. We also learnt the concept of Conditional Probability

Patient → ~~W~~, ~~H~~, $P(\text{diab}) = \dfrac{\#\text{diabetic}}{\#\text{patients}}$

$= 0.25$

W → 120 kgs $\quad P(\text{diab}) \longrightarrow 0.25 \uparrow$

$P(\text{diab} \mid W = 120\text{kg}) = 0.6$

$P(\text{diabetic}) \geq 0.5$

D ND

Probability & Statistics : Predict D → D/ND , w.h

$D \searrow 1$, $ND \searrow 0$

w.h - input

$P(y=1 \mid w.h)$ - Output

$P(y=1 \mid W=\_\_, h=\_\_)$

Will it remain same if you get to know that the person with high weight is extremely tall?

$P(y_i) \xrightarrow{\text{Thinking}} y_i$

Data perspective . .

For ith patient

ith sample

- Consider:

  height as first feature, **f1 = xi1**

  weight as the second feature, **f2 = xi2**

- Predict if the patient is diabetic, **yi**

X_i

$[x_{i1}, x_{i2}] \xrightarrow{ML} y_i$

Probability & Statistics : "Predictive"

In ML, we can think it like this

- Given feature of ith patient, xi=[xi1, xi1 , .. , xid]
- You have to predict yi

Q. But how is it different from what we have been doing?

Q. We have calculated probability before also, right?

Case 1: Data : Calculations

$$\frac{\#\ \text{favorable}}{\#\ \text{total}}$$

DATA WAS THERE

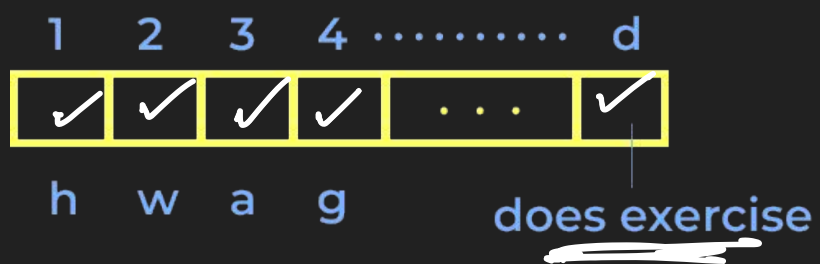
Case 2: X is not present in data

PREDICTION

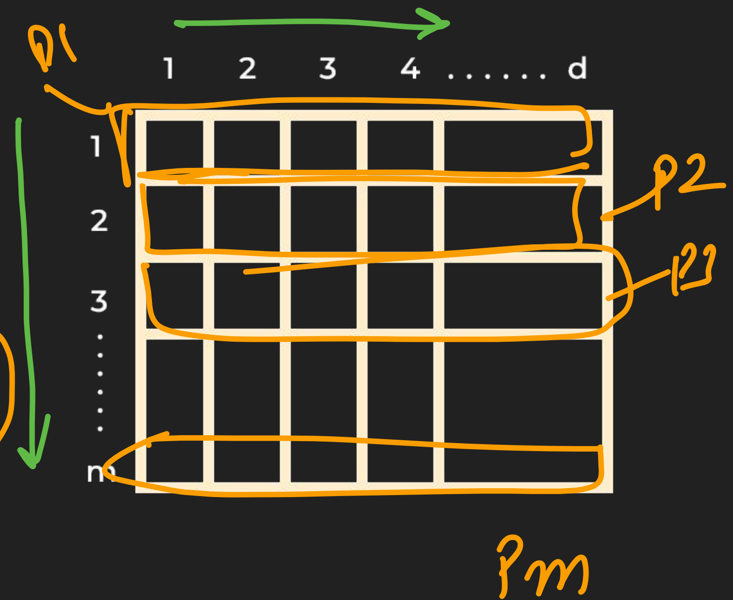
Linear Algebra : D/ND

You have some data X with features - height, weight, age, gender...x_{d}

- One sample d-dimensional vector



- m samples of patients



m x d

Maths: Matrix of m d-dim vectors. R^d

Numpy: 2D array (2-dimensional)

POINTS TO REMEMBER

- We use numpy and pandas for data loading, processing and manipulation
- The type of plot can be decided looking at the kind of data we have. # variable, Types.
- In conditional probability, we check the probability of an event with respect to another event which has happened.
- If xi is present in the data, it becomes a data analysis /calculation problem, else it becomes a ML problem.

  Predict ka nature

ML v/s Classical Programming:

Definition by Arthur Samuel -

ML provides ability to learn without being explicitly programmed.

Q.1 - But can a "dumb" computer perform anything by itself if don't tell what to do?

Q.2 - If somehow, let's assume that it is possible, shouldn't we use ML for everything?

Lets understand with a task

Given an email, you have to identify whether is spam or not spam

Example: Email → Spam vs Not Spam

Classification.

Google

How does the SDE pipeline look ?

Historical Data

E1	E2	E3	E4	------	Em
↓	↓	↓	↓		↓
S	NS	S	S		NS = labels

Look for interesting patterns (if-elif else)

Purchase/Buy/Delete Immediately

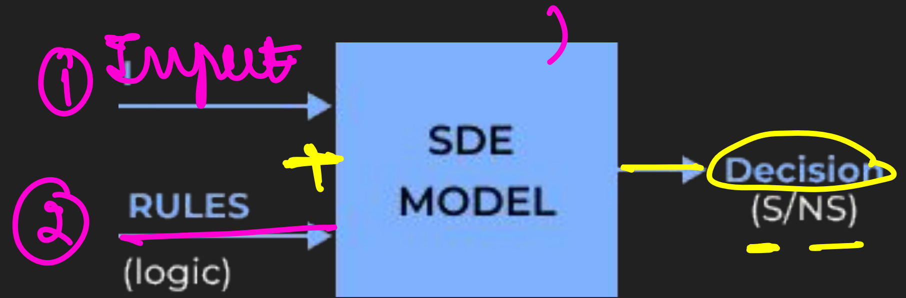
S

NS

long if-elif-else ladder.

We can visualise the SDE approach as

1. Inputs - Input text, Rules (written by programmer)
2. Output - Decision (Spam/Not Spam)



Discovered by an Domain expert

$\text{Text} \xrightarrow{\text{NLTK}} \text{Features (BOW, n-gram etc)}$

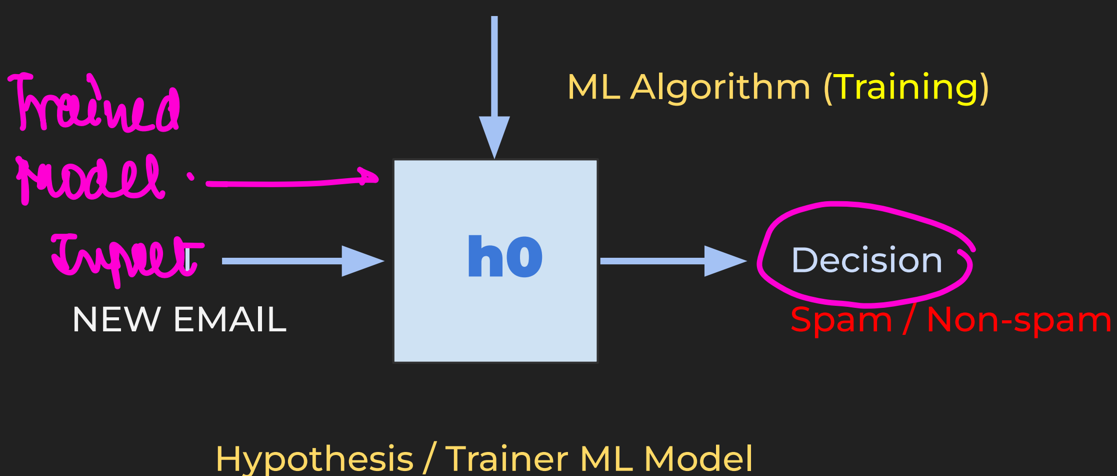
$\Downarrow$

LOOK FOR MEANINGFUL PATTERNS/RULES

Instead of finding and coding them, you code "How to find rules"

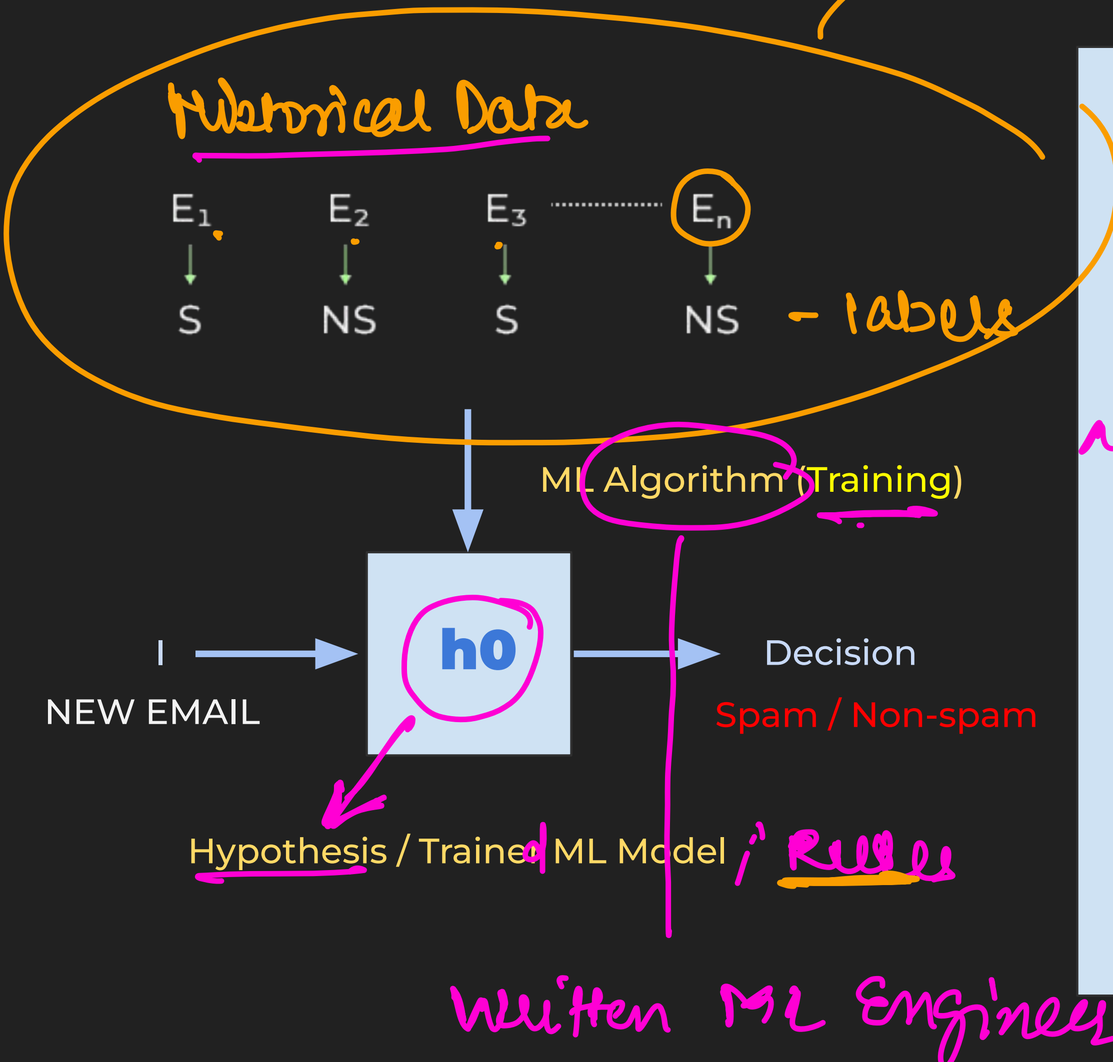
ML Coding → Code which can find rules/patterns by itself (using data)

MODEL TRAINING to look for patterns.



How does the ML pipeline look ?

Use this for model training



We can visualise the ML approach as

1. **Inputs** - Training Data (Text, Decision (Spam/Not Spam))
2. **Learning Algorithm**
3. **Output** - Trained Model (also called Hypothesis)

Given a "unknown" sample, we can use hypothesis to "predict" whether its a spam or not spam

Tom Mitchell's Definition

He uses three terms to define it

Elements of statistical learning

TASK	EXPERIENCE	PERFORMANCE

Let's take an example :

Task (T) To be able to ride bicycle

To classify email as S/NS

Experience (E) Someone teaching you how to pedal

– Training by showing historical emails labels

Performance (P) How well you are riding bicycle

Accuracy of classification.

- **Experience (E)** - You show the some training data to "train" a system using some "algorithm"
- **Performance (P)** - You must define a quantitative measure of how the "model" is performing

We ask, **has the model learnt** to perform better at the **task of classification**

Tom Mitchell's Definition -

Study of algorithms which:

- **improves performance (P)**
- **At task (T)**
- **With experience(E)**

ETP Example:

T	E	P
CASE I Predict price of a stock.	Using historical price data for training a model which can output some "Real Values"	Any metric which is big if difference in P_true and P_predicted is big, else small
CASE II Segment Customers	Using transactional data for training a model which can put similar customers into same cluster	Any metric which maximises the inter cluster difference and minimises the intra cluster distance
CASE III Categorise Image into orange, kinno, mossambi	Train using Images + Labels using a model which can "classify" an image as O, K, M	Any metric which gives a high(er) score for correct label, and low(er) score for incorrect labels

POINTS TO REMEMBER

- SDE pipeline : Provide Rules and Input, Model will give decision.
- ML Pipeline : We provide Input data and ML algorithm, algorithm will mine rules itself and decide output based on them
- Tom-Mitchell def : Study of algorithms which improves performance (P) at Task (T) with Experience (E)

E T P definition

ETP

Based on the type of "task" ML Algos can be categorised as

Classification:

Data → C1, C2, C3 — classifying into one of the category

Fruit - OK/NO | DND / B-NB

Regression:

Predict a real/continuous value

Stock Price Predict

REAL

Clustering:

Group similar samples

Customer segmentation → S1, S2, S3, S4

Predefined ?

Recommendation: (LATER)

M1 M2 M3 M4 M5

M9 M1 M2 M10 M12 — similar

M9

Forecasting: (LATER)

X Y Z A B C D ??

D1 D2 D3 D4 D5 D6 D7

Past prices

Price prediction ??

ETP

Based on the type of "learning" use in training the model

Training a model with supervision

Supervised Learning:
Training data has labels also

Unsupervised Learning:
Training data do not have labels.

Reinforcement Learning:
LATER

supervised

Case1) 100 images O,K,M with labels

Case2) 100 Image - distinguish.

unsupervised