Recap

ML → Data   learning models
     & algo

     ↳
     (Ingredients) ————————————— Training

Target Variable
given ← → not given
  ↓   ↓
Supervised   Unsupervised

Questions to be asked before selecting a ML algo.

Q1 → what is the target variable

Q2 →

Q3 → What is the type of target variable

_____

Supervised Learning →   (Target Variable → given in historical)
                                              data   )

Unsupervised learning →   (Target Variable → not given in historical data)
                                    (Y)

Supervised

Classification   →   Regression

y → Categorical   y → Real
  ↳ numerical   values
     discrete
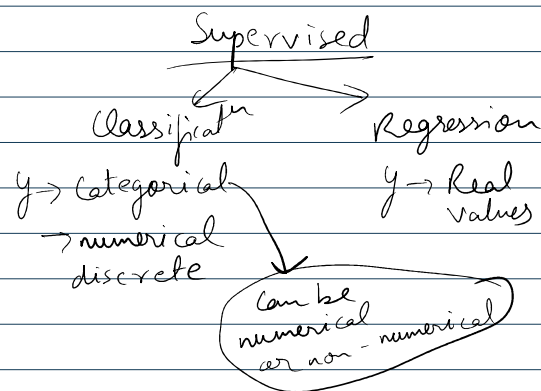
Y ↘
(Data set mai numerical discrete/
          continuous
       ho sakta hai)

(Can be
numerical
or non - numerical)

Problem 0 : Given height & weight of an Individual, predict the gender.

(h_1, w_1) ↗ Male
            or
          ↘ Female

eg

X	Y
5	25
1	1
0.5	0.25
2	4

3 → 9

Algo → Takes the data
           ↓
    Contains I/p & O/p
           ↓
      learns &
      generates a

$$\begin{bmatrix} 0.5 & 0.25 \\ 2 & 4 \end{bmatrix}$$

$$y = X^2$$

learns &
generates a
model which
defines a relat\"ship
b/w I/p & O/p

Whenever you have a prob → Identify the learning
           → Identify the target variable
           → Determine type of " "
           → Finalize the task to be done.
           → Identify the algo. to be used.
           → Business Understanding
           → EDA
           → Data Preparation (Cleaning + transformation)

Model building using ML or DL
         ↓
      Evaluate the model

     Regression   Classificat\"
       ↓   ↓
   →   M S E   ① Accuracy
   →   R M S E   ② Confusion Matrix
   → Mean Absolute   ③ Precision & Recall
     Error   ④ F1 - Score
   → Coeff of determinat\" (R^2) ⑤ ROC - AUC
   → Adjusted R^2 ↓   ↓
            Reciever   Area under
            operator istic   the curve.
            characterise
            curve
            ⑥ log - loss

for Problem ①
     → Historical data → given ✓
     →   Input → Height + Weight
     →   Output → gender
     →   gender → categorical (Male/Female)
        ↳ Target variable
     → Target Variable in Historical data → yes → Supervised learning
     → Task → classification
     → choose an algo → logistic Reg, KNN, SVC, Random Forest C, DTC, etc
     → choose an evaluation matrix

↳ Accuracy, F1-Score, Precision & Recall. etc.

Problem 1: Given height, predict weight

Input → height

O/p → weight → Target Variable. (Continuous Numerical Feature)

Supervised learning → Target is in historical data.

Task: Regression

Algo → linear Reg., KNN Reg., SVR, DTR, GBDT Reg. etc.

Evaluate → MSE, RMSE, R^2, Adjusted R^2

Problem 2 → IRIS dataset

Input → SL, SW, PL, PW

Output → Species → Target variable (Discrete)

Target → Categorical
↳ present in historical data

learning → Supervised

Task → Classification

Algo → logistic Reg, ---- etc

Evaluation → Precision, Accuracy, etc.

Problem 3 → Given an email → spam or not

Input → Body, Title

O/p → spam/ham → Target Variable (Discrete)

Target → Categorical → present in historical data

learning → Supervised

Task → Classification

Algo → KNN, SVC, DTC

Evaluation → F1-Score, Recall, precision etc.

Problem 4 → Given a Text review, predict the Rating

Input → review

O/p → Rating (1,2,3,4,5) → Target Variable

Target Variable → Discrete, numerical → present in historical data

Task → classification

learning → Supervised.

Algo → KNN, SVC, DTC

Evaluation → F1-Score, Recall, precision etc.

Problem 5: Give an Image of handwritten digit, Identify the digit.

      Input → Image of digit

      Output → 0 - 9 → Target Variable

      Target → Discrete, numerical

         ↳ present in historical data.

      Task → classification

      learning → Supervised

      Algo → KNN, SVC, DTC

      Evaluation → Precision, Recall, Accuracy etc.