

Machine Learning (emphasis on Supervised Learning)

T-1

1. What is ML?
2. Types of ML
 - ↳ supervised learning → classification & regression
 - ↳ unsupervised learning → clustering & association rule
 - ↳ reinforcement learning (only intro)
3. Pipeline of ML
4. Some ^{performance} measures for model
 - ↳ accuracy, confusion matrix, FPR (false +ve rate), TPR (true +ve rate), ROC curve and AUC (area under curve), log-loss funcⁿ, R^2 metric
5. Linear and logistic regression
6. Decision trees
 - i) entropy
 - ii) information gain
 - iii) gini index
7. Bias, variance, overfitting, underfitting

What is ML?

By Tom Mitchell

"A machine or an agent is said to learn from experience with respect to some class of tasks and a performance measure 'P' if (the learner's) performance at task in class measured by 'P' improves with experience."

- 1) Answering exams in a particular subject
- 2) Diagnosing patients of a specific illness
- 3) Diagnosing plant disease

1st task is defined (only define those tasks which can be performed by machine)

Second thing we need is a kind of performance measure.

m → no. of marks you are scoring
↓
performance measure

Third important component is experience. With experience the performance has to improve.

eg-i) If you are giving same mock test again and again, Then definitely your performance (marks) will increase.

ii) More patients a doctor will diagnose more experienced he will be or in other words diagnosing performance will increase.

Three components of ML

- 1) you need a class of task
- 2) you need a performance measure
- 3) you need some well defined experience

this is nothing but data or dataset

Note: The kind of learning where you are learning to improve the performance based on experience is called inductive learning.

Different ML Paradigms

1) Supervised learning

Learn an i/p to o/p mapping

eg - job post

↳ multiple comments

+ve

-ve

either +ve or -ve

Dataset

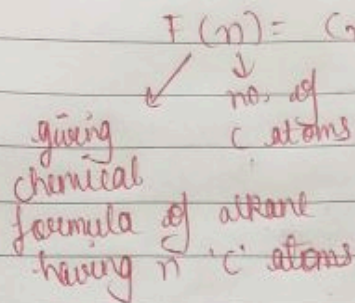
Comment	class (label)
x_1	y_1
x_2	y_2
x_3	y_3
x_n	y_n

i/p to o/p mapping

we will try to make a function out of this data

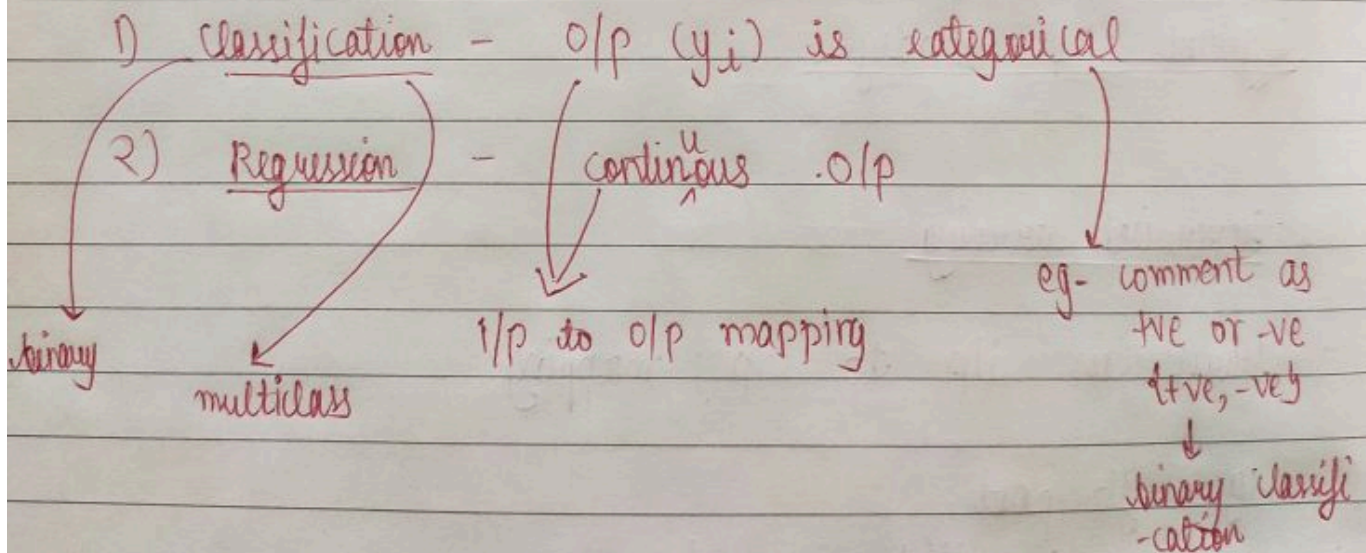
$$y_i = f(x_i) \rightarrow \text{or, model}$$

eg- In chemistry we have general expression or formula for alkanes



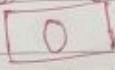
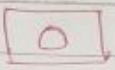
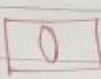
Suppose we have calculated $y_i = f(x_i)$ using dataset. Then, a new comment x_q is given to us. We need to classify it as +ve or -ve. Or, in other words we have to get y_q using $y_i = f(x_i)$.

Supervised learning



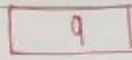
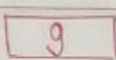
MNIST dataset (contains handwritten images of digits from '0' to '9')

eg- images of zero

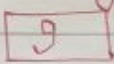
  

images of 1

Suppose user gives me a new image containing a digit.

eg- 

O/P set: {0, 1, 2, 3, ..., 9}

→ multiclass classification

We need to identify this digit given by user.

Regression Model

Tomorrow it will rain (1) or not (0) is a classification problem.

But, amount of rain (generally measured in cm) is regression problem.

Consider a dataset

Index	Nationality	Weight	Hair color	Height
x_1	-	-	-	-
x_2	-	-	-	-
\vdots				
x_n	-	-	-	-

Using given dataset we need to predict height of a person x_q with following attribute value

(x_q , India, 68, black, ?)

Height may lie in a range.

Eg- 120cm - 185cm

There are infinite no. in given range.
