6/8/24 ★★

Supervised ML (has annotation)

Classification
(predicted output is discrete output)

Regression
(only kind of supervised ML output is pointing the spectrum (not a discrete output))

Annotations obtained by → direct manual work (human)

→ multi annotators req (Inter annotation score & annotator matrix taken)

→ Arrnot always can be obtained from natural world itself (Ex → temp)

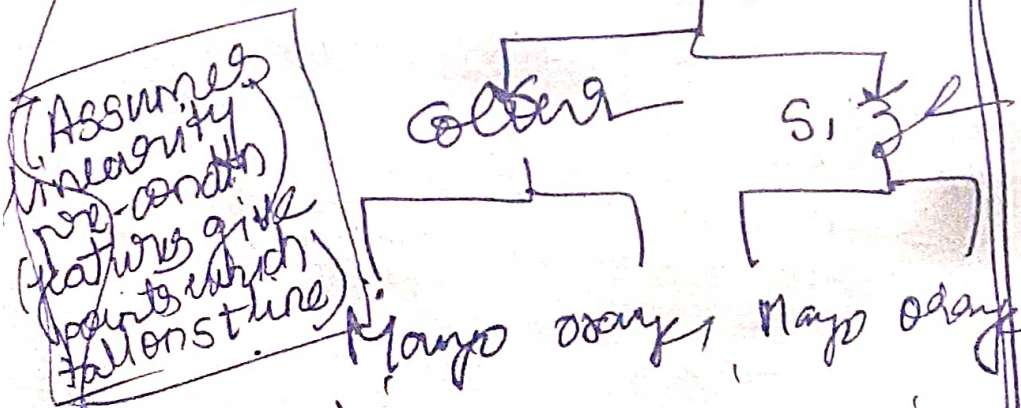
* best algorithm ≠ powerful algorithm
* can be determined by explainability / wt
* Variance is degree of dispersion in training data.
  - If variance is less, model becomes overconfident
  - If " is more, model doesn't converge
  - Here variance tradeoff req.

6/8/24 classification.

(i) K-Nearst Neighbour → easiest way to do ML. high performance
  - no training phase at all.
  - closest label to the req example
  - const time complexity 1

(ii) Decision trees :
  - ~~Find best possible division~~ : Find useful features to divide continously

Ex :



(Assumes linearity, re-creates features give points which fall on straight line)

(iii) (Naive) Bayesian classifiers :- (Find labels from given features)
  - Assumes features are independent

Support Vector Machine → based on Kernel (Non linear relationship) assumption
→ not as complex as neural network
→ applicable for only a binary network SVM
(ex: cat & dog)
Use cases

Regression Models → Linear regression
→ Logistic regression (Non linear)

16/8/24

Unsupervised ML (No annotations)

(i) Clustering Algorithms → to find out no of clusters in a data
→ Guesswork of user req (parametric algorithms)
→ Non-parametric algorithm dont req guesswork (Bayesian)
[ex: Chinese restaurants, Indian buffets analogy]
(Used in anomaly detection)

(ii) Topic Modelling → statistical model
→ finds statistical distribution to users find overlapping themes.

Send it to my

Generative Adversial Network → Rae b/n 2 parts of
Neural Network models
resulting in synthetic datapoints
which look real.