Machine learning

→ Machine learning - field of study that gives computer ability to learn without being explicitly programmed

→ Traditional Programing: Rules → [ ] → Answers; Data → [ ]

→ Machine learning: Answer → [ ] → Rules; Data → [ ]

→ why machine learning
- do not require human experts
- Black box human expertise (face, handwriting recognition)
- Rapidly changing phenomenon (fraud detection)
- Need for customisation (Personalisation (recommendation)

→ Machine learning Process flow

Training data → learn Algorithm → Build model → Perform → feedback → learn Algorithm

→ learning vs design -
- designer can help create data without noise leading to accurate machine learning model.
- design helps machine learning gather better data.
- design help set expectation and establish trust with user.
- designers specifically - UX can add clarity to ml powered interfaces.
- Machine learning customises interface to user needs.

→ Error - It measures how wrong was our estimation.

→ Noise - It refers to irrelevant information.

→ Training vs testing
- Training set - dataset that is feed to ML model so that it learn patterns and trends.
- testing set - Once model is trained, we can make prediction on testing set.
- Validation set - Training set is divided into, train & validation set based on validation result, model is trained.

Machine learning use case
- Financial service (Fraud, Risk detection)
- Healthcare (Disease prediction)
- Agriculture
- Ecommerce (Recommendation)
- Travel

Steps in ml

Import data → Prepare, clean data → fit model → Evaluation new data → Accuracy.

→ Data acquisition - Involves collecting and acquiring data from various sources. like. census data, logs of server etc.

→ Data preparation - The data collected is not clean there are some error which needs to be cleaned.

→ Hypothesis and modelling - based on requirement a model is created using dataset.

→ Evaluation - Model is evaluated on test dataset.

→ Deployment - In this model is deployed in market

→ operation & optimisation - Retraining of model.

Types of learning
- Supervised (labelled data)
- Unsupervised (unlabelled data)
- Reinforcement (Rewardbased learning)

→ Supervised learning - Regression (Numerical prediction of continous value)
- Classification (prediction of categorical value)

Regression - linear regression, multiple linear regression

classification - Decision tree, Random forest, Naive bayes, SVM

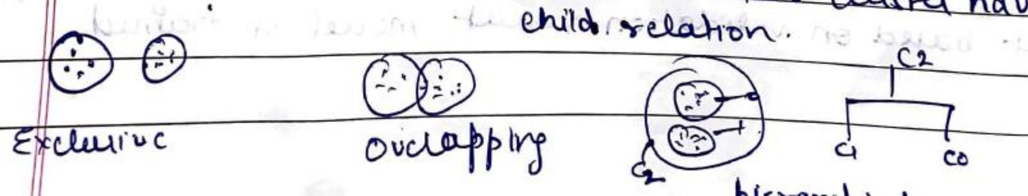
→ Unsupervised learning - clustering - (finding groups)
- Density estimation
- visualisation.

clustering - means grouping objects based on similarity.

Ex - kmeans clustering, c-means clustering, Hierarchical clustering

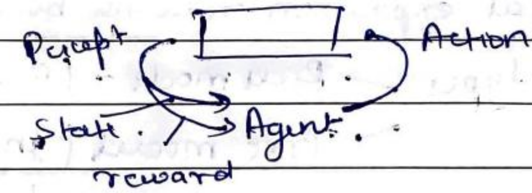
use - marketing, insurance, search engine, seismic zones.

Types - Exclusive clustering - Independent clusters, no overlapping.
- overlapping clustering - an item can belong to one or more cluster
- Hierarchical " → when two cluster have parent child relation.



computer

- Reinforcement learning - In this "agent" take action in environment in order to maximise the reward.

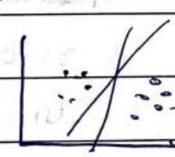


- Learning models - Geometric model
Probabilistic model
logical model

Geometric model - These model define similarity by considering the geometry of the instance space. Here feature could be points in two dimension (x and y axis) or in 3-Dimension.
Geometric model are of two type - linear model
Distance based model.

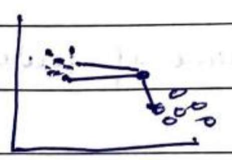
→ linear model - In linear models the equation used is.

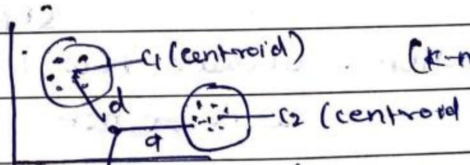
$y(x) = a+bx$



Example - hour study vs marks
rainfall vs yield.

→ Distance based model - In this model concept of distance is used for classification. Some algo based on these are -
→ Nearest neighbour classifier. (Distance can be euclidean, manhattan etc.)
→ k means clustering





New point goes to one having small distance

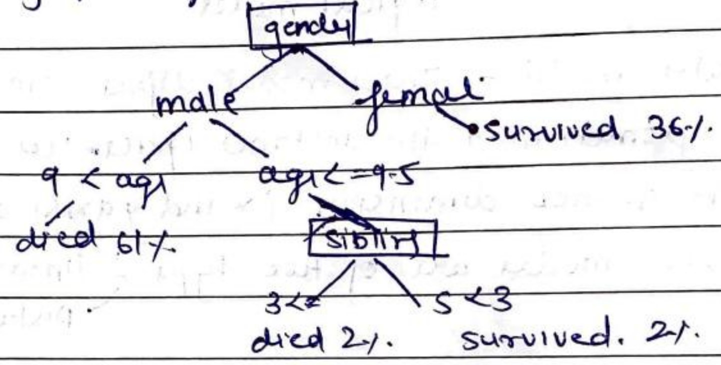
→ Probabilistic model - A probabilistic model is based on theory of probability. This modelling represent and manipulate the level of uncertainity.
Example - Naive bayes classifier. - It is based on the Naive bayes conditional Probability formula.

$$P(A/B) = \frac{P(B/A)\,(A)}{P(B)}$$

→ logical learning models - logical model uses a logical expression to divide instance space into segments and construct grouping. A logical expression returns boolean value.
There are two types — Rule model - (Based on if then rule)
— Tree model (In this tree structure is formed).

Example of logical model - decision tree algo.

<u>survival of passenger on titanic</u> -



→ Features - In training data we have matrix where each row is vector and column is dimension we call each dimension a feature.

- Feature selection - In very high dimensional data such as DNA and text document we need some import feature, selecting those feature is called feature selection.
- Methods — univariate → Pearson correlation, f-score, chi-square
  — multivariate - dimensionality reduction, SVM
- limitation - Unclear how to tell in advance if feature will work or not
  - how many feature to select.

<u>Unit 2</u> - classification and <u>regression</u>

- Regression - numerical prediction of continous value.
- classification - categorical prediction.
- <u>Linear regression</u> - It is supervised machine learning algorithm in this we predict y (dependent variable) based on given independent variable (x).

function → $y = \theta_1 + \theta_2 \cdot x$

θ_1 - intercept θ_2 - coefficient of x.

once we get bee θ_1 & θ_2, we get best fit line. which is then used for further predictions.

$$y = \theta_1 + \theta_2 x$$

$$\theta_2 = \frac{\Sigma (x-\bar{x}) * (y-\bar{y})}{\Sigma (x-\bar{x})^2} \qquad \bar{x} \text{ - mean.}$$

θ_1 - use - $\bar{y} = \theta_1 + \theta_2 \bar{x}$

- find best fit line for given data
-

	x	y	$x-\bar{x}$	$y-\bar{y}$
	95	85	17	8
	85	95	7	18
	80	70	2	-7
	70	65	-8	-12
	60	70	-18	-7
	390	385		
Avg	78	77		

put in formula -

$\theta_2 = 0.644$

now: eq

$\bar{y} = \theta_1 + \theta_2 \bar{x}$

$77 = \theta_1 + 0.644 \times 78$

$\theta_1 = 26.78$

line eq^n = $26.76 + 0.644x$

- multiple linear regression - It involves more than one predictor variable.

  eq^n = $y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 - - -$

- Non-linear regression - In this data is not linearly dependent. so we need more accurate model. this can be achieved by non-linear regression.

  $$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

- exponential regression - if there is constant rate change we use linear model but in constant percent rate of change exponential model is good.

  $$f(t) = a * b^{(t)}$$

→ logistic regression - It is a supervised classification algorithm similar to linear regression but target value is discrete. It uses logistic function for classification of class or category.

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}}$$ logistic func^n.

linear regression	logistic regression
Predicts continous value	predict categorical value

→ Model performance

Model performance are metrics used to evaluate performance of machine learning model. These help in finding reliability and accuracy of model.

different metrics are-

- Accuracy - It is fraction of correctly classified sample out of total number of samples.
- Precision - Precision is fraction of true positive prediction out of all positive prediction. It predicts how many positive prediction made are actually true.
- Recall - It is fraction of true positive prediction out of all actual positive cases. It detect how well the model is able to detect positive case.
- F1score - It is harmonic mean of precision and recall. It provides a single metric that balances precision & recall
- Confusion-matrix - A confusion matrix is table that summarises performance of binary classifier.

~~Roc curve~~

- Mean absolute Error - Average of absolute difference between predicted and actual value.
- Mean squared error - It is average squared difference b/w predicted and actual value.

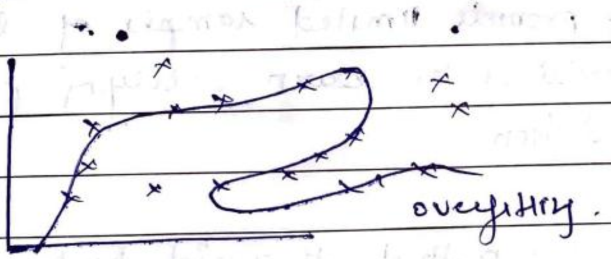
confusion matrix

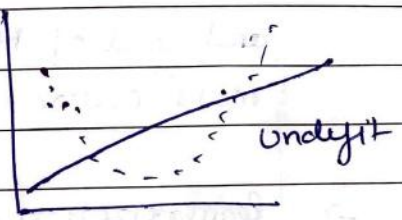
	Predicted no	Predicted yes
Actual No	TN	FP
Actual Yes	FN	TP

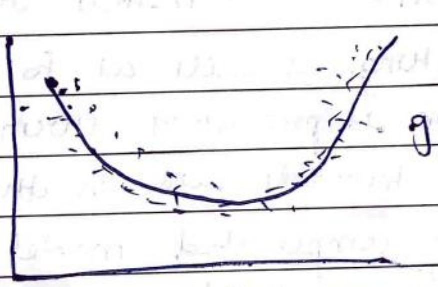
Precision = $\frac{\Sigma\, TP}{TP+FP}$ Recall - $\frac{TP}{TP+FN}$ F1score = $\frac{2}{\frac{1}{recall}+\frac{1}{Precision}}$

→ Overfitting - overfitting occurs when a model is too complex and has learned noise in the data instead of underlying relationship. As a result model performs well on training data but poorly on test data.

→ Underfitting - It occurs when model is too simple and cannot capture underlying relationship. As a result model poorly performs on both training and test data.







→ Bias - It refers to difference between predicted value and true value. A model with high bias consistently predict same value this is because model is too simple and do not capture relationship in data. High bias lead to underfit and model is not able to fit to training data.

→ Variance - It refers to variability of model's prediction for a given input. A model with high variance is sensitive to small fluctuation and produce very different prediction. It lead too overfitting when model become too complex. and fits well on train set and poor on test set.

Reduce bias	Reduce variance
- Increase model complexity	simplify data
- Add more feature	feature selection
- Increase train data	Early stopping
- Regularization	Regularization.

A good model have low bias and low variance.

→ Generalization generalization is the model's ability to give sensible output to sets of input that has never seen before. generalization examine how well new ~~data~~ model can predict new data.

→ A generalise model neither underfit or overfit.

→ The theory of generalization is based on idea that the training data only provide limited samples of distribution and goal of ML model is to learn underlying pattern to make accurate prediction.

→ Regularization - It is a method to avoid high variance and overfitting as well as to increase generalization. Regularization is process of shrinking or regularizing coefficient towards zero in this way it prevent learning of complicated model.

- It enchances generalization capabilities.
- Poor generalization is becuz of overfitting or underfitting
- for poor generalization we apply regularization.

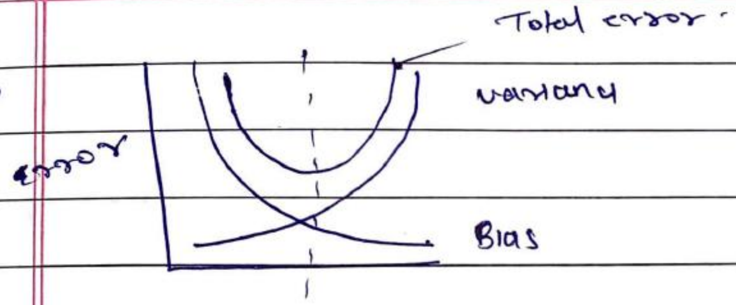
→ L1 Regularization also called lasso regularization It adds a penalty term to the loss function that is proportional to ~~square~~ absolute value of coefficient.

L1 - $cost = \sum_{i=0}^{N} (y_i - \sum_{j=0}^{M} x_{ij} w_j)^2 + \lambda \sum_{j=0}^{M} |w_j|$

→ L2 Regularization, called ridge regularization, it add penalty term to loss function that is proportional to square of coefficient.

$$cost = \sum_{i=0}^{N} (y_i - \sum_{j=0}^{M} x_{ij} w_j)^2 + \lambda \sum_{j=0}^{M} w_j^2$$

→ L3 Regularization - very less used in this penalty term to loss function is proportional to cube of coefficient.

→ 
Total error
variancy
error
Bias