Date: 12/04/23

Machine Learning

Machine Learning: Machine learning is like teaching a computer to do things on its own without being explicitly programmed. Just like how you learn from your experiences, a computer can learn from data.

Types of Machine Learning:

1. Supervised Learning
2. UnSupervised Learning
3. Reinforcement Learning

1. Supervised Learning: - filtering spam mails

The machine learns from training data that is labeled. In labeled data, the output is already known. The model just needs to maps the input to the respective outputs.

- Most Used Algorithms:
  - Linear Regression
  - Logistic Regression
  - Support vector Machine
  - K Nearest Neighbor
  - Decision Tree
  - Random Forest
  - Naive Bayes

- Application:
  They are generally used for Solving classification and regression problems.

Date:

Flipcart uses to figure out products that are suited for you.

2. Unsupervised Learning:

It uses unlabeled data to train machines. Unlabeled data doesn't have fixed output variables. The model learns from the data, discovers the patterns and features in the data, and returns the output.

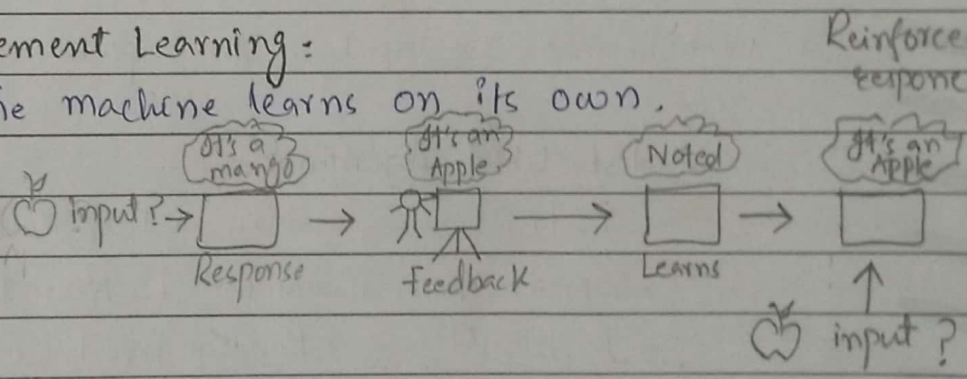
- Most Used Algorithms:
  - K Means Clustering
  - Hierarchical clustering
  - DBSCAN
  - Principal Component Analysis.

- Application:

They are generally used for solving clustering and association Problems.

3. Reinforcement Learning:

The machine learns on its own.



- Algorithms:
  - Q learning
  - Sarsa
  - Monte Carlo
  - Deep Q Network

- Application:

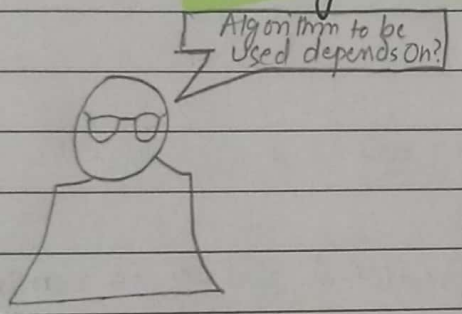
They are widely used in gaming industries to build games. It is also used to train robots to do human tasks.

Date:

Supervised Vs Unsupervised Learning

Supervised	Unsupervised
• Labeled Data	Non labeled Data
• Direct feedback	No feedback
• Predict Output	Find hidden structure in data.

The Right Machine Learning Solution?



- The Problem Statement
- The Size, Quality, and nature of the data.
- Complexity of the algorithms.

Methods to Solve Particular Problems

Classification

Used when the output is Categorical like 'Yes' or 'No'

Algorithms Used

- Decision Tree
- Naive Bayes
- Random Forest
- Logistic Regression
- KNN

Clustering

Used when the data needs to be organized to find Patterns in the case of

- ~~Product recommendation~~.
- Algorithm Used
  - K Means

Regression

Used when a value needs to be Predicted like the ~~stock Prices~~

- Algorithm Used
  - Linear Regression

Date:

Linear Regression

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and a single output variables (y)

For Single LR:

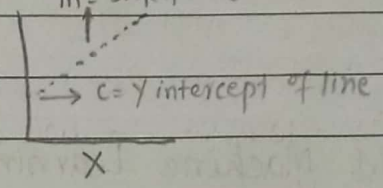
$$Y = mx + C$$

∵ m = slope of the line

C = y-intercept (value of y when x is zero)

$$m = \frac{\Sigma(x-\bar{x})(y-\bar{y}i)}{\Sigma(x-\bar{x})^2}$$

$$C = y - mx$$



For multi LR:

$$Y = C + m1x1 + m2x2 + \cdots + mnxn$$

We use Linear Regression where Predicted value is contineous.

e.g 1. Home Prices 2. weather 3. Stock Prices

Applications

1. Sales Forecasting: To Predict future Sales based on the Past Sales data and other variables such as marketing Spend or pricing.
2. Stock market Analysis: To analyze the relationship between different Stock Prices and make Prediction about future prices.
3. Medical Research: To analyze the relationship between different medical variables, such as blood pressure and heart rate.
4. Marketing Analysis: To analyze the relation between different marketing variables, Such as advertising Spend and Sales.
5. Climate Modelling: To model the relationship between different Climate variables, Such as temperature and Percipitation.

Date:

Logistic Regression

Logistic regression is one of the techniques use for classification.

Classification: Where Predicted value is categorical.

e.g: 1 Will customer by life insurance? Yes/NO, true/false, 0/1

2 Which Party a Person is going to vote for?

1. Democratic 2. Republican 3. Independent.

Classification Types:

1. Binary: To classify input examples one of two classes or categories.

e.g # 1

2. Multiclass classification: To classify input examples one of more than two classes or categories.

e.g # 2

Sigmoid function: It is used for binary classification. in logistic regression and other ML algorithms.
It is commonly used in LR to model the probability of an example belonging to Particular class.

$$\text{Sigmoid}(z) = \frac{1}{1+e^{-z}}$$

∵ e is Euler's number
e = 2.71828

The transformation result in an S-Shaped curve with output value ranging from 0 to 1

$$y = mx + b \longrightarrow y = \frac{1}{1+e^{-(mx+b)}}$$

Logistic Regression

Date:

Logistic Regression Algorithm Steps:

1. Collect a dataset consisting of independent variables and a binary dependent variables.
2. Split the dataset into training and testing dataset.
3. Fit a logistic curve to the training data using maximum likelihood estimation, which finds the parameters that maximize the likelihood of the observed data given the model.
4. Evaluate the performance of the model using the testing data, which can be done by calculating matrics such as accuracy, Precision, recall, F1-score.
5. Use the trained model to make Predictions for new data points.

Applications:

1. Medical research: To Predict the Probability of a disease based on different medical variables, such as, age, gender and BP.
2. Marketing analysis: To predict the probability of a Customer buying a Product based on different marketing variables such as price, promotion, and Product features.
3. Credit scoring: To predict the probability of a borrower defaulting on a loan based on different variables such as income, credit score, and loan amount.
4. Sentiment analysis: To predict the probability of a text being positive or negative on different features such as word frequency and Sentiment analysis.

Date:

Gradient Descent and Cost Function

Gradient descent is an optimization algorithm that is used to find the minimum of a cost function in a Supervised learning Problem.

The cost function is a mathematical function that measures the difference between the predicted output and the true output for a given set of input variables.

The goal of gradient descent is to minimize this cost function by adjusting the parameters of the function that we are trying to learn.

Application :

1. Linear regression: To find the parameters of a linear regression model that minimize the mean squared error between the predicted and true outputs.

2. Logistic regression: To find the parameters of a logistic regression model the minimize the cross-entropy loss between the predicted and true outputs.

3. Support vector machines: To find the optimal separating hyperplane between two classes of data.

4. Neural Networks: To train the weights and biases of a neural network model by minimizing the cost function.

Date:_____

Decision Tree

Decision Tree is a Supervised ML algorithm used for both regression and classification tasks. It creates a tree-like model of decisions and their possible consequences.

- Each node in the tree represents a feature.
- Each branch represents a decision based on that feature.
- The leaves of the tree represent the outcomes or Predicted values.

Working:

The algorithm works recursively splitting the data based on the most significant feature, i.e, the feature that provides most information gain.

∴ Information gain is the measure of the reduction in entropy or impurity acheived by splitting the data based on particular feature.

The Process of selecting the most significant feature and splitting the data is repeated untill the tree is fully grown, or a stopping criterion is met (such as maximum depth or minimum number of samples required at each leaf node)

Once the tree is built, it can be used to predict the class or value of new data by the following path of decision down the tree untill a leaf node is reached, which provides the predicted outcomes.

Date:

Decision Tree Cont:

Applications:

① Classification: To classify data into different categories. For example, it can be used to classify customers into different Segments based on the behaviour.

② Regression: To predict a continuous output value. For example, it can be used to Predict the price of a house based on its features.

③ Anomaly detection: To identify anomalies in the data, i.e, data Points that do not fit the general Pattern of the data.

④ Feature Selection: To identify most important feature in a dataset, which can be useful for feature selection in other machine learning models.

Date:_____

K Means Clustering Algorithm

It is an unsupervised machine learning algorithm used for grouping similar data points into clusters.

- The algorithm works by iteratively Partioning data Points into clusters based on their proximity to a set of K centroids.
- The centroids are calculated as the mean of all data points assigned to a given cluster.
- The algorithm then iterates untill the centroids no longer change significantly, or untill a maximum no: of iterations is reached.

Applications including:

1. Customer Segmentation: To group customers based on their buying Patterns, demographics, or other attributes.

2. Image Segmentation: To group Pixels in an image into different Segments based on their color or intensity values.

3. Text clustering: To group similar documents or text Snippets. ~~based on their context~~.

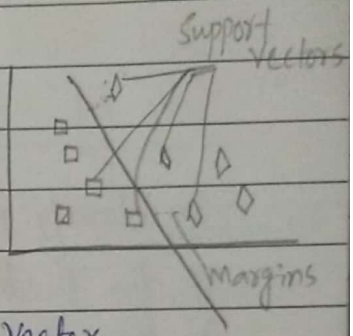
4. Anomaly detection: To identify data Points that are significantly different from the rest of the dataset. Useful to detect outliers or anomalies.

5. Recommender Systems: To group similar Products or items based on user preferences, which can be used to create Personalized recommendations.

Date:

Support Vector Machine

SVM is a Supervised machine learning algorithm used for classification and regression tasks.
It works by finding a hyperplane that seperates the data into different classes while maximizing the margin between the hyperplane and the nearest datapoints of each class.
The hyperplane is defined by a set of weights, which are learned during the training process.



- In SVM, the data points that are closed to the hyperplane are called Support vectors, and they play a crucial role in defining the hyperplane.
- The distance between the hyperplane and Support vector is called the margin, and SVM tries to maximize margin to reduce the risk of overfitting and improve the generalization performance.

Applications:

1. Classification: To classify data into different categories. i.e, it can be used to classify images of handwritten digits into different numbers.
2. Regression: To Predict a continuous output value. i.e, it can be used to Predict the Price of a house based on its features.
3. Anomaly detection: To identify anomalies in the data. ie, data Points that do not fit the general Pattern of the data.
4. Text classification: SVM can be used for text classification tasks such as Sentiment analysis, Spam filtering, and topic classification.

* However, it may not Perform well on non-linearly seperable data. which may requiere more complex variation of the algorithm such as
  * Kernal SVM. $z = x^2 + y^2$

Date:

Random Forest Algorithm

It is a popular machine learning algorithm used for both classification and regression tasks.
It is an ensemble learning that construct multiple decision trees and combine their predictions to make the final Prediction.

* Think of it like a group of friends who each have different ideas about what to do, but they discuss and compromise to come up with a fun plan that everyone can agree on.

- In Classifications tasks, this is done by taking the majority votes of the predicted classes.
- In Regression tasks, the final Prediction is the average of the Predicted values.

Applications:

1. Classification: To classify objects or events into different categories based on their features.
2. Regression: To Predict a continuous value. i.e, Price of a house.
3. Feature Selection: To identify the most important feature that contribute to the outcome, which can be useful in FS and dimensionality reduction task.
4. Anomaly detection: To identify unsual or anomalous observation in dataset.
5. Recommendation System: To build recommendation system that suggests Products or services based on user Preferences and historical data.
6. Image Processing: To object recognition and image segmentation.

Date:

Naive Bayes

It is a Probabilistic ML algorithm that is used for classification tacks. It is based on bayes theorm, which describe the probability of a hypothesis given the available evidence

In Naive Bayes, the presence of one feature does not affect the Probability of the presence of another feature.

- Types of Naive Bayes Algorithm:

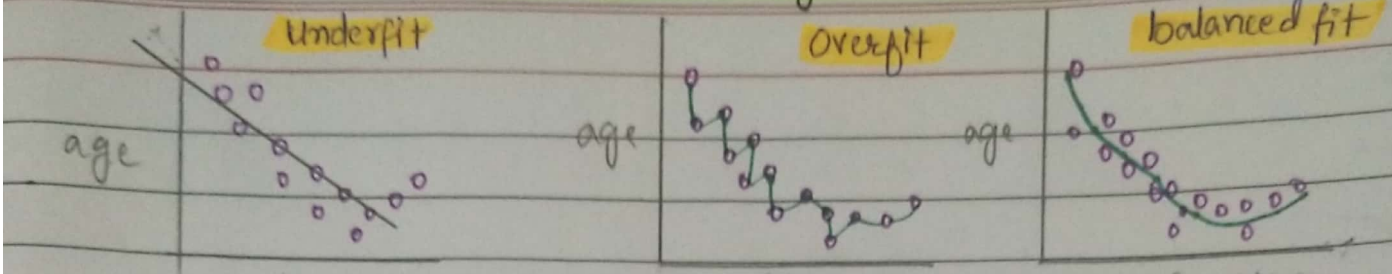
1. Gaussian Naive Bayes: This algorithm assumes that the numerical features in the dataset follow a Gaussian or normal distribution. It is used in cases where all our features are continuous.
2. Multinomial Naive Bayes: It is used for text classification Problems where the data is represented as word counts or frequency of occurrence of words. It assumes features are discrete and follow multinomial distribution
3. Bernoulli Naive Bayes: This algorithm is similar to Multinomial Naive Bayes but is used when the feature are binary or boolean. It follows Bernoulli distribution.

- Application

1. Text classification: Spam filtering, Sentiment analysis, topic classification In above tasks, algorithm learns to classify text documents based on their content
2. Recommendation System: To build RS that suggests Products or Services to user based on their symptoms and medical history.
3. Medical Diagnosis: It is used to predict the probability of a Patient having a Particular disease based on their symptoms and Medical History.

Date:

L1 and L2 Regularization



matchwon = $\theta_0 + \theta_1 \times age$

match won = $\theta_0 + \theta_1 \times age + \theta_2 \times age^2 + \theta_3 \times age^3$

matchwon = $\theta_0 + \theta_1 \times age + \theta_2 \times age^2$

Overfitting: It only Performs well on training data but Performs poorly on testing data. It occurrence simply means that our model or the algorithm does not fit the data well enough. high var, low Bias

Underfitting: Underfitting results in poor Performance on both training and the test data. low var, high Bias

Balanced fitting: Balanced fitting results in gotd Performance on both training data and testing data. Low Var and Low Bias

:. To achieve balanced fiting, it is important to choose appropriate model complexity, use regularization techniques such as L1 or L2 regularization, and use validation techniques such as Cross Validation to assess the performance of the model on new data.
Moreover can use dimensionality Reduction and Ensemble Techniques such as Bagging, Boosting.

Date:

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

Mean Squared Error

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{predicted})^2$$

OR

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x_i))^2$$

(x_i: features)

L2 (Ridge): L2 adds a penalty term to the cost function that forces the model to keep the magnitude of all the features' coefficients small.

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x_i))^2 + \lambda \sum_{i=1}^{n}\theta_i^2$$

(λ: Tunning nob)

L1 (Lasso): L1 adds a penalty term to the cost function that forces the model to reduce the magnitude of less important features' coefficients to zero.

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x_i))^2 + \lambda \sum_{i=1}^{n}|\theta_i|$$

Date:

K-nearest neighbors (KNN)

KNN is a classification algorithm that determines the class of a data point by finding the K nearest data points to it, based on a similarity metric such as Euclidean distance. The class of the data point is then determined by a majority vote among the K-nearest neighbors.

∵ Euclidean distance : It is defined as distance between two points.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

One of the strengths of KNN is its ability to handle non-linear decision boundries, making it a good choice for complex classifications problems. It can be used for both binary and multi-class classification problems.

∴ It can be also sensitive to the choice of K, which can affect its performance.

Date:

Principal Component Analysis

It is a Process of figuring out most important features or Principal components that has most impact on the target variable.

* Few things to Keep in mind before using PCA.

1. Sclale Features before Applying PCA
2. Accuracy might drop

PCA is called dimensionality reduction technique as it can help us reduce dimensions.

∴ What if we get rid of non important feature?

1. Faster training and intervence
2. Data visualization becomes easier.

Date:

Dummy Variables & One HotEncoding

Dummy Variables

```
pd.get_dummies(df.town)
```

(df.town: Dataframe; column name)

One Hot Encoding

```
From SKlearn.preprocessing import LabelEncoder
le = labelEncoder()        # class object

df.town = le.fit_transform(df.town)
```

(df: Dataframe; town: column-name)

One Hot EnCoder

```
from SKlearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder(categorical_features = [0])
X = ohe.fit_transform(x).toarray()
```

(categorical_features = [0]: column; x: independent data frame)

Date:

Cross Validation Techniques

1. **KFold**: It is a simple cross-validation technique that randomly splits the data into K equally sized folds and uses each fold for testing once, while the rest are used for training.

2. **Stratified KFold**: It is similar to KFold, but it ensure each fold has same proportion of classes as the entire data-set, which is important when dealing with imbalanced dataset. It is a modified version of KFold.

3. **Cross-Val-Score**: It is a function in SciKit-learn that automates the process of KFold or Stratified KFold cross-validation and calculates evaluation metric (such as accuracy or F1 score) for each Fold. It simplies the process of cross-validation and is useful for quickly comparing different model or hyperparameters.

Date:

Hyper-Parameter Tuning

Hyperparameter tuning is the process of selecting the best set of hyperparameters for a ML algorithm to achieve optimal performance on a given dataset.
It helps to avoid overfitting or underfitting of the model and can improve the performance of generlized data.

It can be done manually by selecting different values for each hyper parameter and testing the model performance, or it can be automated using techniques like GridSearchCV or RandomizedSearchCV.

GridSearchCV: Grid Search CV evaluates all possible combinations of hyperparameters with a specified range, and selects combinations with best performance based on a specified evaluation metric.
It can be computationally expensive, especially when dealing with a large number of hyperparameters.

RandomizedSearchCV: It randomly samples hyperparameters from a specified distribution, and evaluates the performance of each combination to select the best performing hyperparameters.
It can be computationally more efficient, espeially when dealing with large number of hyperparameters.

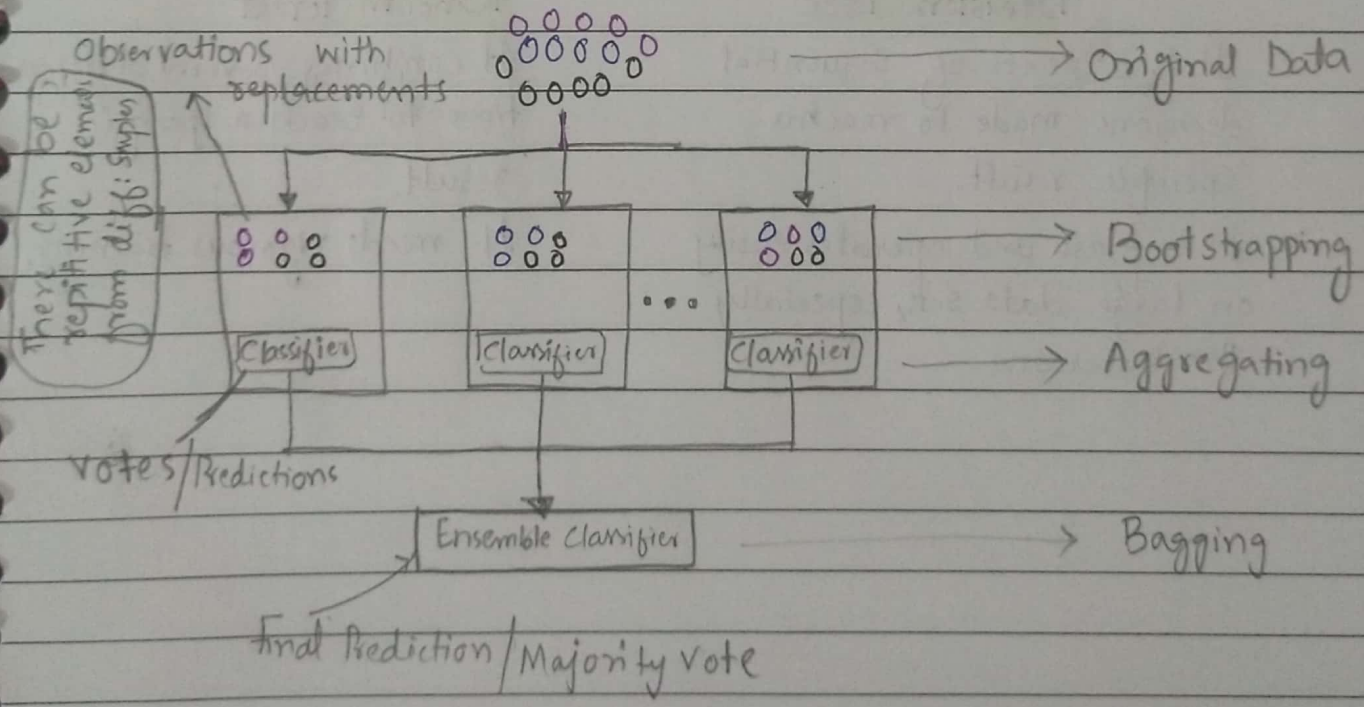
Date:

Ensemble - Bagging

Ensemble learning is a technique in machine Learning where multiple models are combined to make more accurate Prediction than any individual model alone.
It's like a team of experts working together to solve a Problem.

Bagging, short for Bootstrap Aggregating, is an ensemble learning technique where multiple models are trained on different subset of training data. Bagging used in statistical classification and regression. It decreases the variance and helps to avoid overfitting. Each model in the ensemble, known as a "base model" or "weak learner".

Implementation:



Date:

Differences

Random Forest vs Bagging

	Random Forest	Bagging
1.	It is an ensemble Technique	It is also an ensemble Tech:
2.	It uses random feature selection at each split	It does not consider random subset of feature at each split

Decision Tree vs Random Forest

	Decision Tree	Random Forest
1.	It is a series of sequential decisions made to reach a specific result.	It combines several decision trees to reach a specific result.
2.	It is fast and operates easily on large data sets, especially the linear one.	It needs rigorous training.