① Can you name and explain four of the main challenges in machine learning?

Ⓐ We have Five type's of main challenge's in machine learning.

① Not enough training data:

Let's say for a child, to make him learn what an apple, all it takes for you to point to an apple and say apple repeatedly.

For a simple task, it needs thousands of examples to make something out of it. and for advanced tasks like image (or) speach recognition, it may need lakh(millions) of examples.

② Poor Quality of data:

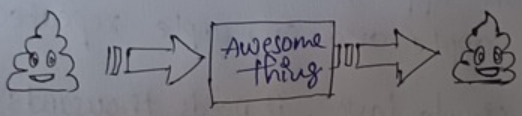
obviously, if your training data has lot of errors, outliers, and noise, it will make it impossible for your machine learning model, to detect a proper underlying pattern. Hence, it will not perform well.

also put in every ounce of effort in cleaning up your training data.

(*) Most Data Scientists spend a significant part of their time in cleaning data".

(*) If you see some of the instance are clear outliers just discard them (or) fix them manually.

(3) Irrelevant Features:-



Garbage in, garbage out (GIGO)".

In the above image, we can see that even if our model is "AWESOME", and we feed it with garbage data, the relevant the data that result will also be garbage (output). our training data data must always contain more, relevent and less none irrelevant feature.

The Credit for a successful meachine project goes to coming up with a good set of features on which it has been trained.

(4) Non representative training data;

To make sure that our model generalizes well, we have to make sure that our training data should be representive.

If train our model by using a non-representative training set, it won't be accurate in predictions it will be biased against one class (or) group.

② List and Explain Risk statistics.

(A) Risk stastics are used in meaching learning to evaluate the performance and generalization of a model some commonly used risk stastics stics include.

① Training error:-

This is the error rate of the model on the training data. The training error is used to evaluate how well model fits the training data.

② Testing error:-

This is the error rate of the model on the test data. The data error (Test) is used to evaluate how ell The model generalizes to new unseen data

③ Cross Validation error:-

This error rate is the model on a validation set that is created by partitioning data into multiple sets.

Cross validation is used to estimate the generalization performance model and to prevent overfitting.

(4) Bias :-

This difference between the expected value of the predictions made by the model and true value of the target variable Bias measures how well the model captures the true relationship b/w the input features and target variable.

(5) Variance :-

This is variability of the models predictions for different training sets, variance measures how sensitive the model is to small changes in training.

(6) mean sure error (MSE) :-

This is average of square difference b/w predicted value and true value. MSE is used to Evaluate overall performance of the model.

(7) Root Mean squared Error :-

This is square root of MSE, RMSE is used to measure average magnitude of error made by model.

In summary risk stastistics are used to evaluate the performance and generalization of ML model. By using multiple risk stastistics, we can gain a more comprehensive understanding of the models strengths and weaknesses.

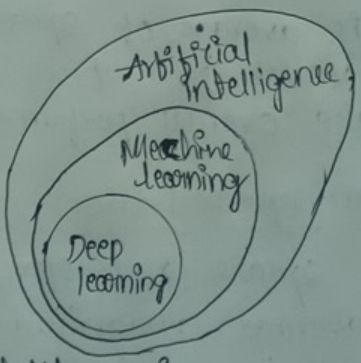
③ Write the differences between Artificial intelligence, machine learning and deep learning.

(A) Deep learning, machine learning, and Artificial intelligence are the most used terms on the internet for IT folks.

However, all these three technologies are connected with each other.

Artificial intelligence (AI) can be understood as an umbrella that consists of both machine learning and deep learning.

We can say deep learning and machine learning both are subsets of artificial intelligence.



Artificial Intelligence:-

Artificial intelligence is defined as a field of science and engineering that deals with making intelligence machine (or) computer to perform human-like activities.

advanced tasks like image
it may need lakh (millions)

④ Bias :-

This difference between the expected value of the predictions made by the model and true value of the target variable Bias measures how well the model captures the true relationship b/w the input features and target variable.

⑤ Variance :-

This is variability of the model's predictions for different training sets, variance measures how sensitive the model is to small changes in training.

⑥ Mean Sure Error (MSE) :-

This is average of square difference b/w predicted value and true value. MSE is used to Evaluate overall performance of the model.

⑦ Root Mean squared Error :-

This is square root of MSE, RMSE is used to measure average magnitude of error made by model.

In summary risk stastistics are used to evaluate the performance and generalization of ML model. By using multiple risk stastistics, we can gain amore comprehensive understanding of the models strengths and weaknesses.

it may need

② Poor Quality of data :-

obviously, if your training data

* MR. John Macarthy is known as the godfather of amazing invention

* AI is defined as the capability of meachine to imitate Intelligent human behavior.

Meachine learning:-

Meachine learning is the concept that a computer program can learn and adapt to new data without human inventions. Meachine learning is a field of artificial intelligence (AI) that keeps a computer's build-in algorithms current regardless of changes in the worldwide economy.

Deep learning :-

Deep learning is the defineed as the subset of meachine learning and artificial intelligence that is based on artificial neural networks. In deep learning, the deep learning, the deep word refers to the number of layers in neural network.

The main difference b|w the meachine learning and deep learning is of presentation of data.

Inthile deep learning technologies is of uses neural networks for learning models.

Assignment - II

① Explain about ANOVA in detail

(ans) ANOVA is useful in scenarios where we need to analyse the impact of categorical features on numerical target variables,

(*) It identify the imporatant feature's in dataset.

(*) check if different groups influence target variable outcomes.

(*) Handling categorical variables in regression models.

ANOVA formula

ANOVA is based on partitioning the total variance to components.

$$F = \frac{\text{Between-group variance}}{\text{Within-group mean}}.$$

Type's of ANOVA in Machine learning

(a) one-way ANOVA

Used when comparing one categorical independent variable with a continous dependent variable.

(b) Two-way variable:-

used when analyzing the effect of two categorical independent variable on continous dependent variable.

ANOVA Feature Selection in machine learning

ANOVA is used for feature selection in regression and classification tasks.

Eg:- by using python

```
from sklearn.f_s import f_classif
import numpy as np
X = np.array([[1],[2],[3],[4],[5],[6]])
Y = np.array([10, 15, 12, 18, 20, 16])
F_stat, p_value = f_classif(X, y)
Print("F-statistic:", F_stat)
Print("P-value:", P_value)
```

∴ ANOVA is used for feature selection in machine learning to identify significant variables.

Alternate methods like chi-square and Mutual information are used based on data types.

(Q) Explain about MNIST class classification technique.

(A) The MNIST database stands for - Modified National Institute of standard and Technology database, is a large database of handwritten digits that is commonly used for training various image processing system.

The database is widely used for training and testing in the field of ML.

It is created by "re-mixing" the samples from NIST's origional datasets.

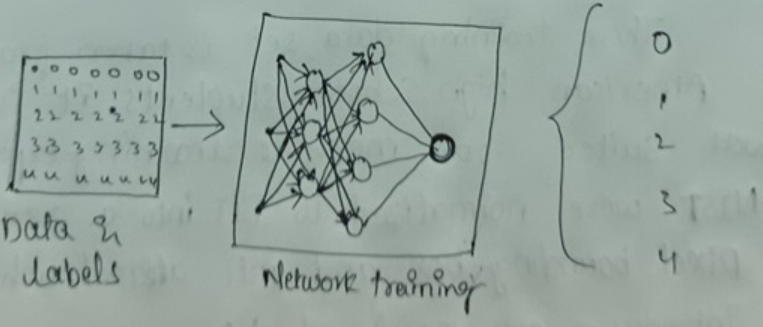
These training data set is taken from American high school students. It is not well suited for mechine learning experiments. NIST were normalized to fit into a 28x28 pixel bounding box and anti-aliased, which introduce grayscale levels.

MNIST data base contains 60000 training images and 10000 testing images.

They use support vector mechine to get an error rate of 0.8%.

Extended MNIST is [EMNIST] is a newer data set developed and released by NIST.

Images in EMNIST were converted into same 28x28 pixel format, but same process, as were the MNIST image.



③ Discuss about multiclass classification techinique.

Ⓐ Multi-class classification is the task of classifying element into different classes. unlike binary, It doesn't restrict it self to any number classification of classess.

Eg:-

- Classification of news in different categories,
- Classifying books, according to the subset.
- Classifying students according to their stream etc.

In there, there are different classess for the response variable to be classified to be in the and thus according to the name, it is multi-class classification.

Popular algorithms are that can be used for multi-class classification.

- K - nearest Neighbours
- Decision Trees
- Naive Bayes
- Random forest
- Gradient Boosting.

⇒ Examples for multiclass classification

- face classification
- Plant species classification
- optical character recognition

∴ Let us suppose we have to do sentiment analysis of a person, if the classes are just "positive" and "negative", then it will be a problem of binary class. But if the classess are "sadness", happiness", "disgusting", "depressed", then it will be called a problem of Multiclass classification.

(Q) Explain KNN algorithm with an example?

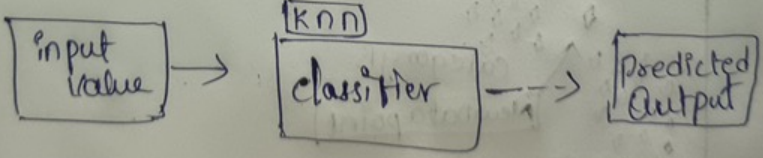
(A) KNN [KNIN] stands for "K- Nearest Neighbour".

(*) It is a supervised meachine learning algorithm.

(*) The algorithm can be used to solve both classification and regression problem statement.

(*) The number of nearest neighbour to a new unknown variable that has to be predicted (or) classified is denoted by the symbol 'k'.

In order to correctly classify the results, we must first determine the value of k.



Step's to working of knn-algorithm.

Step(1):- select the number k of the neighbour

Step(2):- Calculate the Euclidean distance

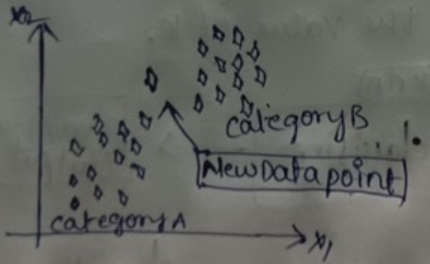
of k-numbers of neighbours.

Step③:- Take the k-nearest neighbour as per the calculated Euclidean distance.

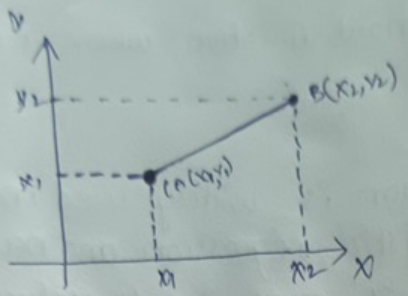
Step④:- Among these k-neighbors, count the numbers of the data points in each category.

Step⑤:- Assign the new data points to the category for which the number of neighbour is maximum.

Step⑥:- Our model is ready.

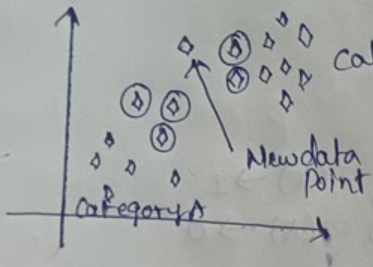


Now you choose K=5 then it Next, will calculate the Euclidean distance b/w the data points.



Euclidean distance b/w A_1 & $B_2 = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

By calculating Euclidean distance we got ~~neigh~~ nearest neighbours.



Large k-values is good, but it may find some difficulties.

Advantages:-

1. It is simple to implement
2. It is robust to the noisy training data
3. It can be more effective if the training data is large.
4. Always needs to determine the value of k which may be complex some time (dis advantage).

(5) What is the sigmoid function? where it can be used ? Explain?

(A) The sigmoid function is widely used in ML, Particularly in logistic regression and neural networks, because it maps inputs values to range b/w 0 and 1, making it useful for Probability estimation

Sigmoid function formula:-

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

(*) As $x \to +\infty, \sigma(x) \to 1$

(*) As $x \to -\infty, \sigma(x) \to 0$

This sigmoid function used in ML, like -.

Binary Classification:- In logistic regression, it converts linear outputs into probabilities.

Neural Networks:- used as an activation function to introduce non-linearity

Probability Interpretation:- since its output is b/w 0 & 1, it is useful for modeling Probabilities.

Deep learning:-

ReLU and Tanh are preferred due to better performance.

- The sigmoid function is a fundamental activation function in machine learning.
- It play's a crucial role in logistic regression and probabilistics models
- However, for deep neural networks, alternative activation functions like ReLU are preferred due to efficiency.

Sigmoid function is like an old-school calculator - not as fast (or) powerful as modern tools, but still useful in right situations.

(6) Illustrate the stacking mechanism in ensemble techniques.

(A) Stacking is an ensemble learning method that improves model performance by combining multiple base learners through a meta-learner.

It is different from bagging and boosting because it's focus on learning.

Stacking consists of multiples:-

Level 0:-
Multiple diverse models are trained independently on the dataset.
Each model makes predictions on the training data.

Level 1:- The predictions from the base models are used as input features for a new model.

- The meta learner is trained to combine the predictions optimally.
- Common meta-learner include logistic Regression, Random Forest (or) Neural Networks.

Final prediction :-

During testing, the base models generate Predictions.

- The meta-model takes these predictions and make the final decision.

to use stacking :- when boosting or bagging isn't giving enough accuracy improvement.

when computation resources allow for multiple models.

Avoid stacking if :

- The base models are two similar.
- The data set is to small
- Computational resources are limited.

Blending & Multi-layer stacking is the Advanced stacking Variants, Techiniques.

① Blending:-

A simpler form of stacking where base model predictions are blended with a validation dataset.

② Multi layer stacking:-

Stacking multiple layers of base models before the final meta model.

③ Simple stacking:-

This is the basic form of stacking.

① Train multiple base models on the same dataset

② Collect their predictions

③ use the prediction as input to train a meta-model that makes the final predictions.

UNIT - III

① Define Boosting? Explain about Ada boosting techinique.

(A) : boosting is an ensemble learning techinique that improves the performance of weak classifiers by combining multiple models sequentially.

(*) Each new model focuses more on the mistakes of the previous models.

(*) The final prediction is a weighted combination of all weak models, leading to a strong classifiers with better accuracy.

Ada boosting techiniques :-

(1) initialize weights :- Assign equal weights to all training samples.

(2) Train weak Learner :-
Train a weak classifier using the weighted data set.

(3) Calculate Error:-
measure the classifier's weighterror.

(4) Compute classifier weights:- Assign a weight to the weak classifier based on its performance

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-e_t}{e_t}\right)$$

(5) update sample weights:-

(a) Increase weights for misclassified samples.

(b) Decrease weights for correctly classified samples.

(6) Normalize weights:-
Ensure the weights sum to 1.

(7) Repeat for multiple iterations:- Train multiple weak classifiers and combine their outputs.

(8) Final strong classifier:-
Combine weak classifier's using their weights $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

(2) what is Bagging and pasting? Explain its implementation with scikit-learn.

(A) (*) In machine learning, some times multiple Predictors grouped together have a better predictive Performance than anyone of the group alone.

(**) These technique are very popular in competitions and production. They are are ensemble learning.

(***) There are several way's to group models. They differ the training algorithm and data used in each one of them and how also they grouped. we will be talking about two methods called Bagging and pasting.

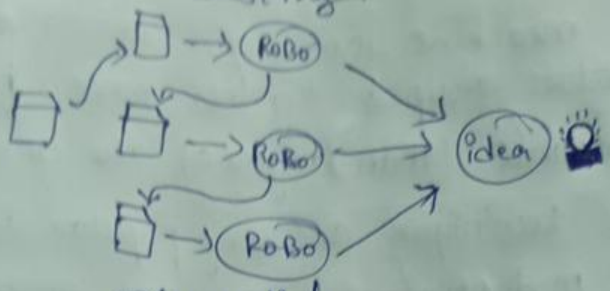
implementation of scikit-learn :-

=) Pip install scikit-learn (installation)

-) load data set :-

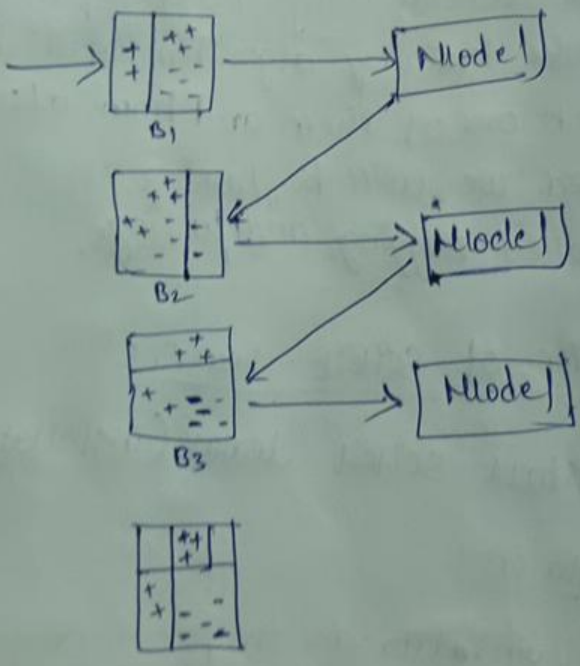
```
from sklearn.datasets import load_iris
import pandas as pd
iris = load_iris()
```

Boosting:- Boosting.



Sequential

Eg:-



⇒ B4 = B1 + B2 + B3

The Ada boost uses Boosting technique.

On ... the final predicti... majority voting.

It doesn't

```
x = iris.data
y = iris.target
df = pd.DataFrame(x, columns= iris.feature_
                                names)
df['target'] = y
Print(df.head())
```

Save and load the model:-

```
import joblib
joblib.dump(model, "decision_tree_
                        model.pkl")
loaded_model = joblib.load("decision
        _tree_model.pkl").
```

(3) What is the difference b/w hard and soft voting classifiers? Explain them.

(A) Hard voting classifier:-

In hard voting, each base model make a prediction, and the final prediction is based on majority voting.

It doesnot consider probability scores, only most frequent predictions.

Sql-

(A) 3 classifiers (A, A, B)

→ The highest majority vote has A got the 'A' as the final prediction.

Features	Hard voting classifier	Soft voting classifier
Decision Method	Majority vote	Weight average of predicted Probabilities
Probability consideration	NO	yes
works Best with	Model with strong individual Performance	Models that provide well-calibrated Probabilities
sensitivity to confidence	No	yes
interpretability	Easier to understand	Slightly complex due to probability averaging
Robustness to Noise	More robust	Can be sensitive to poorly calibrated Probabilities.

1) Can you name and Explain four of the main challenges in machine learning.?

2) List and explain Risk statistics.

3) Explain about MNIST dataset. describe the Procedure to apply classification technique.

4) what is the sigmoid function? where it can be used? ~~Explain~~

5) what is bagging? Explain it's Random forest algorithm.

6) Illustrate the stacking mechanism in ensemble techniques.