1. What Are the Different Types of Machine Learning?

There are three types of machine learning:

1- Supervised Learning

In supervised machine learning, a model makes predictions or decisions based on past or labeled data. Labeled data refers to sets of data that are given tags or labels, and thus made more meaningful.

2- Unsupervised Learning

In unsupervised learning, we don't have labeled data. A model can identify patterns, anomalies, and relationships in the input data

3- Reinforcement Learning

Using reinforcement learning, the model can learn based on the rewards it received for its previous action.

2. What is Overfitting, and How Can You Avoid It?

The Overfitting is a situation that occurs when a model learns the training set too well, taking up random fluctuations in the training data as concepts. These impact the model's ability to generalize and don't apply to new data.

When a model is given the training data, it shows 100 percent accuracy—technically a slight loss. But, when we use the test data, there may be an error and low eficiency. This condition is known as overfitting.

There are multiple ways of avoiding overfitting, such as:

- Regularization. It involves a cost term for the features involved with the objective function
  Regularization works by adding a penalty or complexity term to the complex model
- Making a simple model. With lesser variables and parameters, the variance can be reduced
- Cross-validation methods like k-folds can also be used

1. How Can You Choose a Classifier Based on a Training Set Data Size?

When the training set is small, a model that has a right bias and low variance seems to work better because they are less likely to overfit.

In machine learning, the phrase "models with low bias and high variance" refers to a type of model behavior in the context of predictive performance.

- **Bias**: Bias refers to the error due to overly simplistic assumptions in the learning algorithm. A model with low bias will be flexible and can capture complex relationships in the training data.
- **Variance**: Variance refers to the model's sensitivity to small fluctuations or noise in the training data. A model with high variance will fit the training data very closely, even capturing the noise, but it may not generalize well to unseen data.

2. Explain the Confusion Matrix with Respect to Machine Learning Algorithms.

A confusion matrix is a table used to evaluate the performance of a classification model by summarizing the counts of different types of predictions:

- True Positive (TP): 25
- False Positive (FP): 10
- True Negative (TN): 50
- False Negative (FN): 15

	Actual Positive	Actual Negative
Predicted Positive	25 (TP)	10 (FP)
Predicted Negative	15 (FN)	50 (TN)

- Accuracy = (TP + TN) / Total
- Precision = TP / (TP + FP)
- Recall (Sensitivity) = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- F1-score = 2 * (Precision * Recall) / (Precision + Recall)

3. What Are the Three Stages of Building a Model in Machine Learning?

The three stages of building a machine learning model are:

- Model Building & training
- Model Testing
- Applying the Model & deploying

4. What is Deep Learning?

The Deep learning is a subset of machine learning that involves systems that think and learn like humans using artificial neural networks. The term 'deep' comes from the fact that you can have several layers of neural networks.

5. What Are the Applications of Supervised Machine Learning in Modern Businesses?

- Email Spam Detection
- Healthcare Diagnosis
- Fraud Detection

6. What is Semi-supervised Machine Learning?

In the case of semi-supervised learning, the training data contains a small amount of labeled data and a large amount of unlabeled data.

7. What Are Unsupervised Machine Learning Techniques?

- Clustering problems involve data to be divided into subsets. These subsets, also called clusters, contain data that are similar to each other. Different clusters reveal different details about the objects, unlike classification or regression.

- In an association problem, we identify patterns of associations between different variables or items.

8. What Is ‘naive’ in the Naive Bayes Classifier?

The classifier is called ‘naive’ because it makes assumptions that may or may not turn out to be correct.

The algorithm assumes that the presence of one feature of a class is not related to the presence of any other feature (absolute independence of features), given the class variable.

9. How Will You Know Which Machine Learning Algorithm to Choose for Your Classification Problem?

- If accuracy is a concern, test different algorithms and cross-validate them
- If the training dataset is small, use models that have low variance and high bias
- If the training dataset is large, use models that have high variance and little bias

Bias and Variance usually will have inverse relation. Increasing one will reduce the other.

10. When Will You Use Classification over Regression?

Classification is used when your target is categorical, while regression is used when your target variable is continuous both classification and regression belong to the category of supervised machine learning algorithms.

Examples of classification problems include:

- Predicting yes or no
- Estimating gender
- Breed of an animal
- Type ofcolor

Examples of regression problems include:

- Estimating sales and price of a product
- Predicting the score of a team
- Predicting the amount of rainfall

11. What is a Random Forest?

A 'random forest' is a supervised machine learning algorithm that is generally used for classification problems. It operates by constructing multiple decision trees during the training phase. The random forest chooses the decision of the majority of the trees as the final decision.

12. What is a Decision Tree Classification?

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever-smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

13. Briefly Explain Logistic Regression.

Logistic regression is a classification algorithm used to predict a binary outcome for a given set of independent variables.

The output of logistic regression is either a 0 or 1 with a threshold value of generally 0.5. Any value above 0.5 is considered as 1, and any point below 0.5 is considered as 0.

14. Explain the K Nearest Neighbor Algorithm.

K nearest neighbor algorithm is a classification algorithm that works in a way that a new data point is assigned to a neighboring group to which it is most similar.
In K nearest neighbors, K can be an integer greater than 1. So, for every new data point, we want to classify, we compute to which neighboring group it is closest.

15. What is a Recommendation System?

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system: It's an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

16. What Are Some Methods of Reducing Dimensionality?

You can reduce dimensionality by combining features with feature engineering, removing collinear features, or using algorithmic dimensionality reduction.

Some examples:

1. **Feature Selection**:
    - recursive feature elimination
2. **Feature Extraction (Dimensionality Reduction)**:
    - Principal Component Analysis (PCA
3. **Autoencoders**:
    - Neural network architecture designed to learn efficient encodings of data.

17. What is Principal Component Analysis?

Principal Component Analysis or PCA is a multivariate statistical technique that is used for analyzing quantitative data. The objective of PCA is to reduce higher dimensional data to lower dimensions, remove noise, and extract crucial information such as features and attributes from large amounts of data.

---

18. What do you understand by the F1 score?

The F1 score is a metric that combines both Precision and Recall. It is also the weighted average of precision and recall.

The F1 score can be calculated using the below

formula: $F1 = 2 * (P * R) / (P + R)$

The F1 score is one when both Precision and Recall scores are one.

19. Explain Correlation and Covariance?

Correlation: Correlation tells us how strongly two random variables are related to each other. It takes values between -1 to +1. Formula to calculate Correlation:

Covariance: Covariance tells us the direction of the linear relationship between two random variables. It can take any value between $-\infty$ and $+\infty$.

20. What is Cross-Validation?

Cross-Validation in Machine Learning is a statistical resampling technique that uses different parts of the dataset to train and test a machine learning algorithm on different iterations. The aim of cross-validation is to test the model's ability to predict a new set of data that was not used to train the model. Cross-validation avoids the overfitting of data. K-Fold Cross Validation is the most popular resampling technique that divides the whole dataset into K sets of equal sizes.

1. **Goal**: The goal of k-fold cross-validation is to assess the performance of a machine learning model in a robust way and to avoid overfitting.
2. **Process**:
    - **Step 1**: Split your dataset into k equally sized "folds" or subsets.
    - **Step 2**: Train and evaluate your model k times.
    - **Step 3**: In each iteration, use k-1 folds for training and the remaining fold for validation.
    - **Step 4**: Calculate the average performance across all k iterations.

21. How does the Support Vector Machine algorithm handle self-learning?

SVM is a machine learning algorithm used for classification and regression tasks. It aims to find the best possible line or hyperplane that separates different classes of data points.

14) What is the difference between Type1 and Type2 errors?

Type 1 error is classified as a false positive. I.e. This error claims that something has happened but the fact is nothing has happened. It is like a false fire alarm. The alarm rings but there is no fire.

Type 2 error is classified as a false negative. I.e. This error claims that nothing has happened but the fact is that actually, something happened at the instance.

28) What is your favorite use case for machine learning models?

The decision tree is one of my favorite use cases for machine learning models.

Decision trees are easy to understand and visualize.

Decision trees can capture complex nonlinear relationships in the data

Decision trees can be trained and evaluated quickly

29) Is rotation necessary in PCA?

Yes, rotation is definitely necessary because it maximizes the differences between the variance captured by the components.

30) What happens if the components are not rotated in PCA?

It is a straight effect. If the components are not rotated then it will diminish eventually and one has to use a lot of various components to explain the data set variance.

35) Explain the concept of machine learning and assume that you are explaining this to a 5-year-old baby?

Yes, the question itself is the answer.

Machine learning is exactly the same way how babies do their day-to-day activities, the way they walk or sleep, etc. It is a common fact that babies cannot walk straight away and they fall and then they get up again and then try. This is the same thing when it comes to machine learning, it is all about how the algorithm is working and at the same time redefining every time to make sure the end result is as perfect as possible.

One has to take real-time examples while explaining these questions.

39) What are the different types of algorithm techniques are available in machine learning?

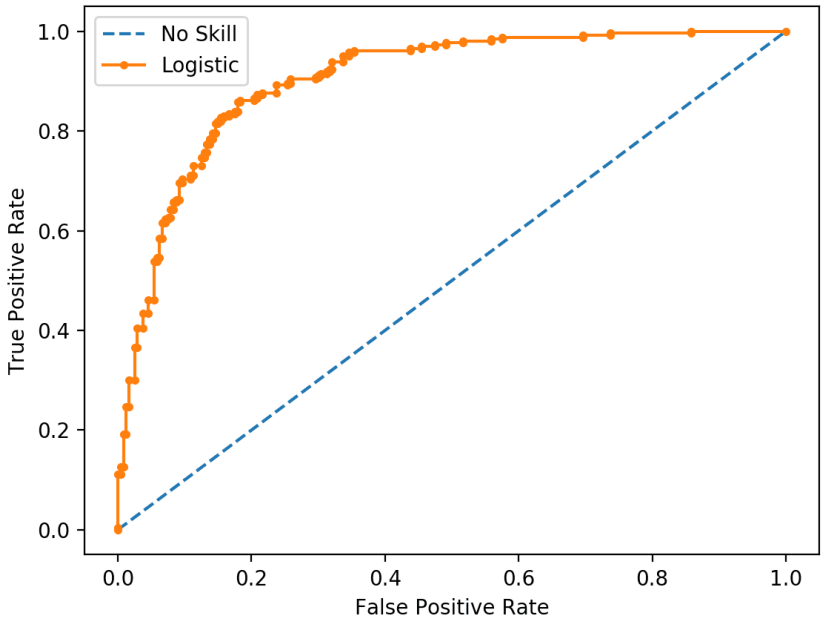
Some of them are :

1. Supervised learning
2. Unsupervised learning
3. Semi-supervised learning
4. Transduction
5. Learning to learn

41) Explain how the ROC curve works?

A ROC curve (receiver operating characteristic) is a graph that shows the performance of a classification model at all classification thresholds. It plots two parameters -

- True positive rate
- False-positive rate



45) What is Bayes Theorem and how it is used in machine learning?

Bayes theorem is a way of calculating conditional probability ie. finding the probability of an event occurring based on the given probability of other events that have already occurred. Mathematically, it is stated as -

$$P(A|B) = \{P(B|A).P(A)\}/P(B)$$

46) What is cross-validation techniques would you be using on a time series dataset?

Cross-validation is used for tuning the hyperparameters and producing measurements of model performance. With the time series data, we can't use the traditional cross-validation technique due to two main reasons which are as follows

- Temporal dependencies
- Arbitrary Choice of Test Set

47) What are the Discriminative and generative models?

To understand these terms better, let us consider an example. Suppose a person has two kids - kid A and kid B. Kid A learns and understands everything in depth whereas Kid B can only learn the differences between what he sees. One day, that person took them to the zoo where they saw a deer and a lion. After coming from the zoo, the person showed them an animal and asked them what it was. Kid A drew the images of both the animals he saw in the zoo. He compared the images and answered "the animal is deer" based on the closest match of the image. As Kid B learns things based on only differences, therefore, he easily answered: "the animal is deer.". In ML, we call Kid A a Generative Model and Kid B a Discriminative Model. To make it more clear, the Generative Model learns the joint probability distribution p(x,y). It predicts the conditional probability using Bayes Theorem. Whereas a Discriminative model predicts the conditional probability distribution p(y|x). Both of these models are used in supervised learning problems.

49) What is more important - model accuracy or model performance?

Accuracy is more important in machine learning models. We can improve model performance by using distributed computing and parallelizing over the scored assets. But accuracy should be built during the model training process.

50) How would you handle an imbalanced dataset?

- Collecting more data to even the imbalances in the dataset.
- Resample the dataset to correct for imbalances.

51) When should you use classification over regression?

, we can say that we would use regression if the outputs are in real numbers and we would go with classifications if the outputs are in the form of classes or categories.

53) What are the data types supported by JSON?

strings, numbers, objects, arrays, booleans, and null values.

25) Which method is frequently used to prevent overfitting?

When there is sufficient data 'Isotonic Regression' is used to prevent an overfitting issue.

27) What is Perceptron in Machine Learning?

In Machine Learning, Perceptron is a supervised learning algorithm for binary classifiers where a binary classifier is a deciding function of whether an input represents a vector or a number.

29) What are Bayesian Networks (BN)?

Bayesian Network is used to represent the graphical model for probability relationship among a set of variables.

30) Why instance based learning algorithm sometimes referred as Lazy learning algorithm?

Instance based learning algorithm is also referred as Lazy learning algorithm as they delay the induction or generalization process until classification is performed.

31) What are the two classification methods that SVM (Support Vector Machine) can handle?

- Combining binary classifiers
- Modifying binary to incorporate multiclass learning

40) What is dimension reduction in Machine Learning?

In Machine Learning and statistics, dimension reduction is the process of reducing the number of random variables under considerations and can be divided into feature selection and feature extraction.

41) What are support vector machines?

Support vector machines are supervised learning algorithms used for classification and regression analysis.

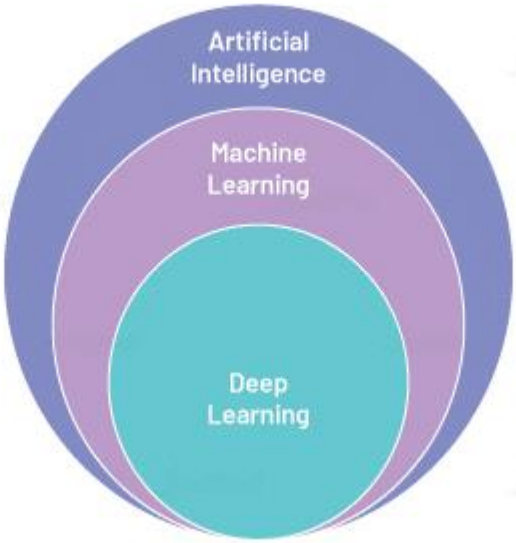
43) What are the different methods for Sequential Supervised Learning?

The different methods to solve Sequential Supervised Learning problems are

- Sliding-window methods
- Recurrent sliding windows
- Hidden Markow models

1. Explain Machine Learning, Artificial Intelligence, and Deep Learning

It is common to get confused between the three in-demand technologies, Machine Learning, Artificial Intelligence, and Deep Learning. These three technologies, though a little different from one another, are interrelated. While Deep Learning is a subset of Machine Learning, Machine Learning is a subset of Artificial Intelligence. Since some terms and techniques may overlap in these technologies, it is easy to get confused among them.



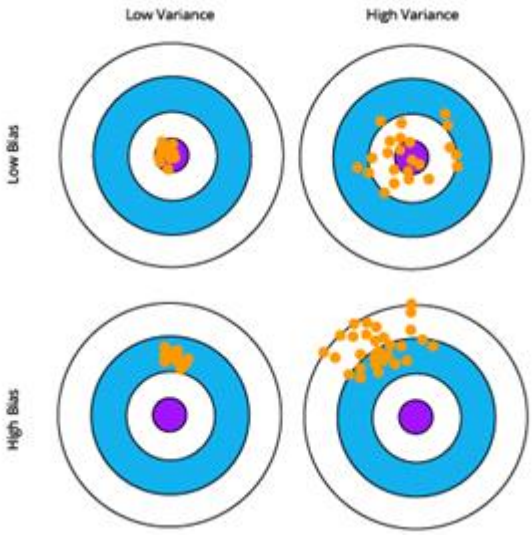
So, let us learn about these technologies in detail:

- Machine Learning: Machine Learning involves various statistical and Deep Learning techniques that allow machines to use their past experiences and get better at performing specific tasks without having to be monitored.
- Artificial Intelligence: Artificial Intelligence uses numerous Machine Learning and Deep Learning techniques that enable computer systems to perform tasks using human-like intelligence with logic and rules. Artificial intelligence is used in every sector hence it is necessary to pursue Artificial Intelligence Course to make your career in AI.
- Deep Learning: Deep Learning comprises several algorithms that enable software to learn from themselves and perform various business tasks including image and speech recognition. Deep Learning is possible when systems expose their multilayered neural networks to large volumes of data for learning.
-

2. What is Bias and Variance in Machine Learning?

- Bias is the difference between the average prediction of a model and the correct value of the model. If the bias value is high, then the prediction of the model is not accurate. Hence, the bias value should be as low as possible to make the desired predictions.
- Variance is the number that gives the difference of prediction over a training set and the anticipated value of other training sets. High variance may lead to large fluctuation in the output. Therefore, a model's output should have low variance.

The following diagram shows the bias-variance trade-off:



3. What is Clustering in Machine Learning?

Clustering is a technique used in unsupervised learning that involves grouping data points. The clustering algorithm can be used with a set of data points. This technique will allow you to classify all data points into their particular groups. The data points that are thrown into the same category have similar features and properties, while the data points that belong to different groups have distinct features and properties. Statistical data analysis can be performed by this method. Let us take a look at three of the most popular and useful clustering algorithms.

- K-means clustering: This algorithm is commonly used when there is data with no specific group or category. K-means clustering allows you to find the hidden patterns in the data, which can be used to classify the data into various groups. The variable *k* is used to represent the number of groups the data is divided into, and the data points are clustered using the similarity of features. Here, the centroids of the clusters are used for labeling new data.
- Mean-shift clustering: The main aim of this algorithm is to update the center-point candidates to be mean and find the center points of all groups. In mean-shift clustering, unlike k-means clustering, the possible number of clusters need not be selected as it can automatically be discovered by the mean shift.

4. What is Linear Regression in Machine Learning?

Linear Regression is a supervised Machine Learning algorithm. It is used to find the linear relationship between the dependent and independent variables for predictive analysis.

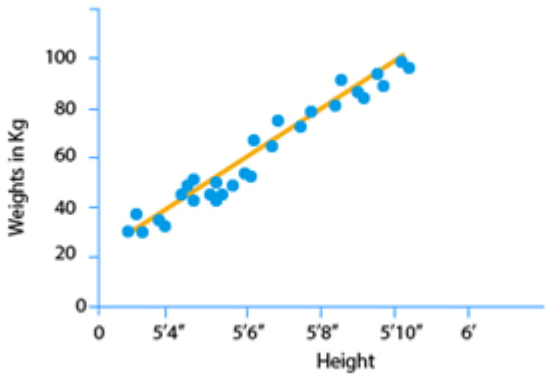
The equation for Linear Regression:

$$Y = A + B.X$$

where:

- X is the input or independent variable
- Y is the output or dependent variable
- a is the intercept, and b is the coefficient of X

Below is the best-fit line that shows the data of weight, Y or the dependent variable, and the



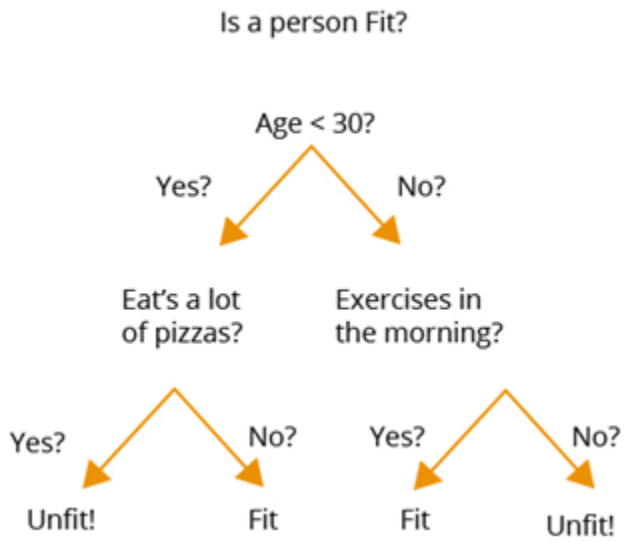
ata of height, X or the independent variable, of 21-year-old candidates scattered over the plot. The straight line shows the best linear relationship that would help in predicting the weight of candidates according to their height.

To get this best-fit line, the best values of a and b should be found. By adjusting the values of a and b, the errors in the prediction of Y can be reduced.

This is how linear regression helps in finding the linear relationship and predicting the output.

5. What is a Decision Tree in Machine Learning?

A decision tree is used to explain the sequence of actions that must be performed to get the desired output. It is a hierarchical diagram that shows the actions.



An algorithm can be created for a decision tree on the basis of the set hierarchy of actions.

6. What is Overfitting in Machine Learning and how can it be avoided?

Overfitting happens when a machine has an inadequate dataset and tries to learn from it. So, overfitting is inversely proportional to the amount of data.

For small databases, overfitting can be bypassed by the cross-validation method. In this approach, a dataset is divided into two sections. These two sections will comprise the testing and training dataset.

7. What is Hypothesis in Machine Learning?

Machine Learning allows the use of available dataset to understand a specific function that maps input to output in the best possible way. This problem is known as function approximation. Here, approximation needs to be used for the unknown target function that maps all plausible observations based on the given problem in the best manner. Hypothesis in Machine learning is a model that helps in approximating the target function and performing the necessary input-to-output mappings. The choice and configuration of algorithms allow defining the space of plausible hypotheses that may be represented by a model.

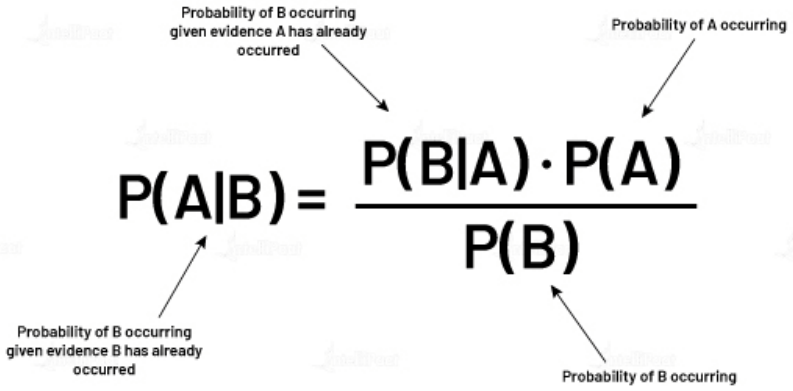
In the hypothesis, lowercase h (h) is used for a specific hypothesis, while uppercase h (H) is used for the hypothesis space that is being searched. Let us briefly understand these notations:

- Hypothesis (h): A hypothesis is a specific model that helps in mapping input to output; the mapping can further be used for evaluation and prediction.
- Hypothesis set (H): Hypothesis set consists of a space of hypotheses that can be used to map inputs to outputs, which can be searched. The general constraints include the choice of problem framing, the model, and the model configuration.

10. What is Bayes’s Theorem in Machine Learning?

Bayes’s theorem offers the probability of any given event to occur using prior knowledge. In mathematical terms, it can be defined as the true positive rate of the given sample condition divided by the sum of the true positive rate of the said condition and the false positive rate of the entire population.

Two of the most significant applications of Bayes’s theorem in Machine Learning are Bayesian optimization and Bayesian belief networks. This theorem is also the foundation behind the Machine Learning brand that involves the Naive Bayes classifier.



11. What is PCA in Machine Learning?

Multidimensional data is at play in the real world. Data visualization and computation become more challenging with the increase in dimensions. In such a scenario, the dimensions of data might have to be reduced to analyze and visualize it easily. This is done by:

- Removing irrelevant dimensions
- Keeping only the most relevant dimensions

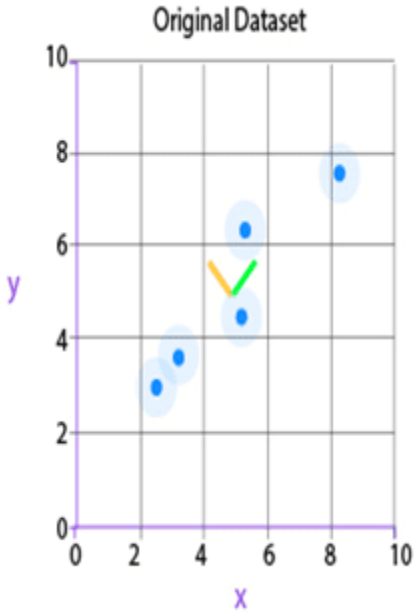
This is where Principal Component Analysis (PCA) is used.

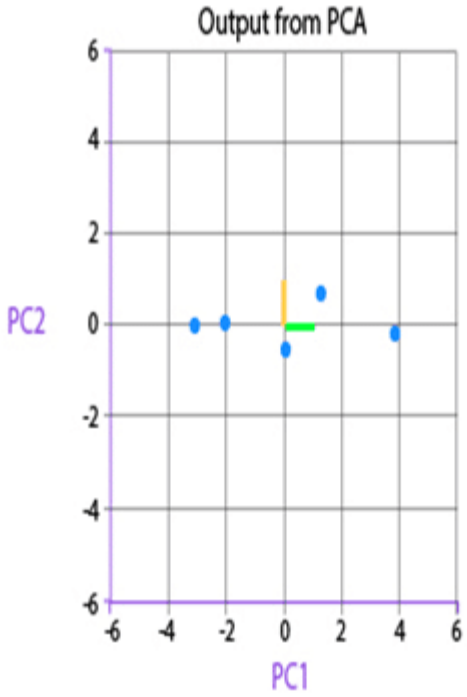
The goal of PCA is to find a fresh collection of uncorrelated dimensions (orthogonal) and rank them on the basis of variance.

Mechanism of PCA:

- Compute the covariance matrix for data objects
- Compute eigenvectors and eigenvalues in descending order
- Select the initial *N* eigenvectors to get new dimensions
- Finally, change the initial n-dimensional data objects into N-dimensions

Example: Below are two graphs showing data points or objects and two directions, one is green and the other is yellow. Graph 2 is arrived at by rotating Graph 1 so that the x-axis and y-axis represent the green and yellow direction respectively.





After the rotation of data points, it can be inferred that the green direction, the x-axis, gives the line that best fits the data points.

Here, two-dimensional data is being represented; but in real life, the data would be multidimensional and complex. So, after recognizing the importance of each direction, the area of dimensional analysis can be reduced by cutting off the less-significant directions.

12. What is Support Vector Machine (SVM) in Machine Learning?

SVM is a Machine Learning algorithm that is majorly used for classification. It is used on top of the high dimensionality of the characteristic vector.

The following is the code for SVM classifier:

```
# Introducing required libraries
from sklearn import datasets
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
# Stacking the Iris dataset
iris = datasets.load_iris()
# A -> features and B -> label
A = iris.data
B = iris.target
# Breaking A and B into train and test data
A_train, A_test, B_train, B_test = train_test_split(A, B, random_state = 0)
# Training a linear SVM classifier
from sklearn.svm import SVC
svm_model_linear = SVC(kernel = 'linear', C = 1).fit(A_train, B_train)
svm_predictions = svm_model_linear.predict(A_test)
# Model accuracy for A_test
accuracy = svm_model_linear.score(A_test, B_test)
# Creating a confusion matrix
cm = confusion_matrix(B_test, svm_predictions)
```

13. What is Cross-validation in Machine Learning?

Cross-validation allows a system to increase the performance of the given Machine Learning algorithm, which is fed a number of sample data from the dataset. This sampling process is done to break the dataset into smaller parts that have the same number of rows, out of which a random part is selected as a test set and the rest of the parts are kept as train sets. Cross-validation consists of the following techniques:

- Holdout method
- K-fold cross-validation
- Stratified k-fold cross-validation
- Leave p-out cross-validation

14. What is Entropy in Machine Learning?

Entropy in Machine Learning measures the randomness in the data that needs to be processed. The more entropy in the given data, the more difficult it becomes to draw any useful conclusion from the data. For example, let us take the flipping of a coin. The result of this act is random as it does not favor heads or tails. Here, the result for any number of tosses cannot be predicted easily as there is no definite relationship between the action of flipping and the possible outcomes.

15. What is Epoch in Machine Learning?

Epoch in Machine Learning is used to indicate the count of passes in a given training dataset where the Machine Learning algorithm has done its job. Generally, when there is a large chunk of data, it is grouped into several batches. All these batches go through the given model, and this process is referred to as iteration. Now, if the batch size comprises the complete training dataset, then the count of iterations is the same as that of epochs.
In case there is more than one batch, d*e=i*b is the formula used, wherein d is the dataset, e is the number of epochs, i is the number of iterations, and b is the batch size.

24. How to handle Missing or Corrupted Data in a Dataset?

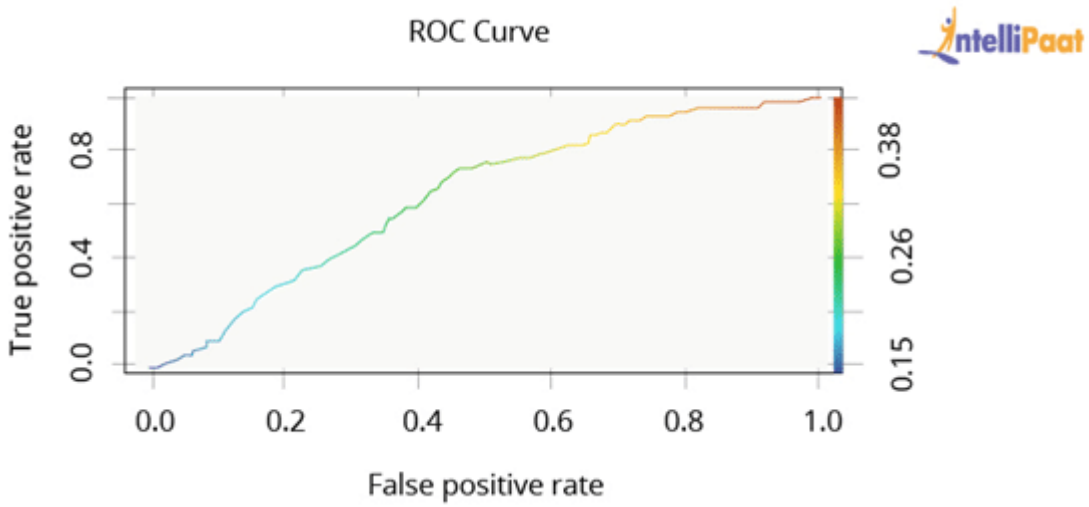
In Python pandas, there are two methods to locate lost or corrupted data and discard those values:

- isNull(): It can be used for detecting the missing values.
- dropna(): It can be used for removing columns or rows with null values.
- fillna() can be used to fill the void values with placeholder values.

26. What is ROC Curve and what does it represent?

ROC stands for receiver operating characteristic. ROC Curve is used to graphically represent the trade-off between true and false-positive rates.

In ROC, the area under the curve (AUC) gives an idea about the accuracy of the model.



The above graph shows a ROC curve. The greater the AUC, the better the performance of the model.

Next, we will be taking a look at Machine Learning interview questions on rescaling, binarizing, and standardizing.

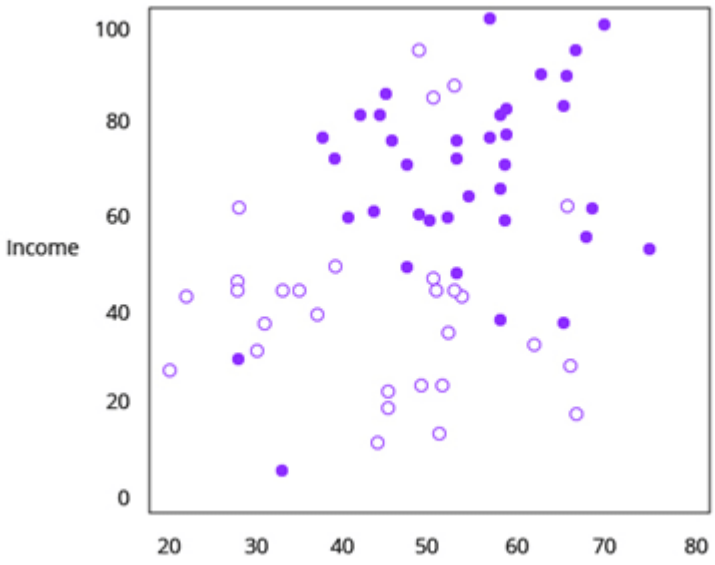
27. Why are Validation and Test Datasets Needed?

Data is split into three different categories while creating a model:

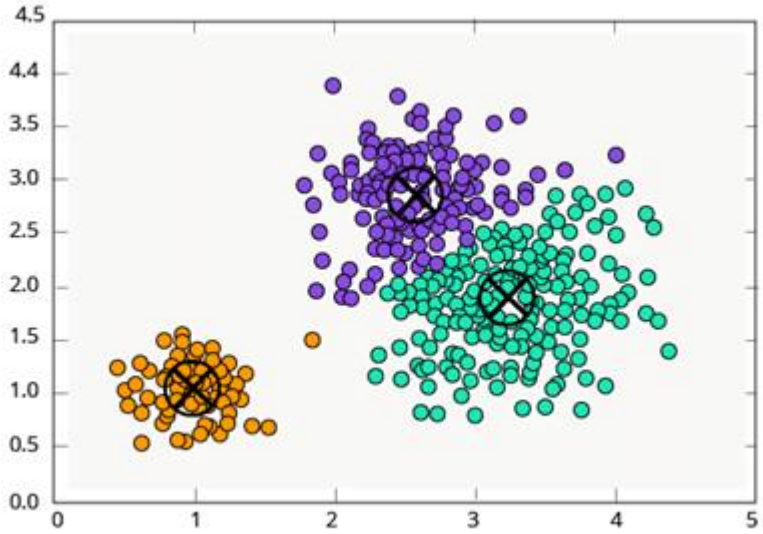
- **Training dataset:** Training dataset is used for building a model and adjusting its variables. The correctness of the model built on the training dataset cannot be relied on as the model might give incorrect outputs after being fed new inputs.
- **Validation dataset:** Validation dataset is used to look into a model's response. After this, the hyperparameters on the basis of the estimated benchmark of the validation dataset data are tuned.When a model's response is evaluated by using the validation dataset, the model is indirectly trained with the validation set. This may lead to the overfitting of the model to specific data. So, this model will not be strong enough to give the desired response to real-world data.
- **Test dataset:** Test dataset is the subset of the actual dataset, which is not yet used to train the model. The model is unaware of this dataset. So, by using the test dataset, the response of the created model can be computed on hidden data. The model's performance is tested on the basis of the test dataset.Note: The model is always exposed to the test dataset after tuning the hyperparameters on top of the validation dataset.

28. Explain the difference between KNN and K-means Clustering

K-nearest neighbors (KNN): It is a supervised Machine Learning algorithm. In KNN, identified or labeled data is given to the model. The model then matches the points based on the distance from the closest points.



K-means clustering: It is an unsupervised Machine Learning algorithm. In K-means clustering, unidentified or unlabeled data is given to the model. The algorithm then creates batches of points based on the average of the distances between distinct points.



31. What is meant by Parametric and Non-parametric Models?

Parametric models refer to the models having a limited number of parameters. In case of parametric models, only the parameter of a model is needed to be known to make predictions regarding the new data.

Non-parametric models do not have any restrictions on the number of parameters, which makes new data predictions more flexible. In case of non-parametric models, the knowledge of model parameters and the state of the data needs to be known to make predictions.

32. Differentiate between Sigmoid and Softmax Functions

Sigmoid and Softmax functions differ based on their usage in Machine Learning task classification. Sigmoid function is used in the case of binary classification, while Softmax function is used in case of multi-classification.

33. In Machine Learning, for how many classes can Logistic Regression be used?

Logistic regression cannot be used for more than two classes. Logistic regression is, by default, a binary classifier. However, in cases where multi-class classification problems need to be solved, the default number of classes can be extended, i.e., multinomial logistic regression.

37. What are the Two Main Types of Filtering in Machine Learning? Explain.

Collaborative filtering refers to a recommender system where the interests of the individual user are matched with preferences of multiple users to predict new content.

Content-based filtering is a recommender system where the focus is only on the preferences of the individual user and not on multiple users.

38. Outlier Values can be Discovered from which Tools?

The various tools that can be used to discover outlier values are scatterplots, boxplots, Z-score, etc.

40. What are the Various Kernels that are present in SVM?

The various kernels that are present in SVM are:

- Linear
- Polynomial
- Radial Basis
- Sigmoid

43. What is Binarizing of Data? How to Binarize?

Converting data into binary values on the basis of threshold values is known as binarizing of data.

44. How to Standardize Data?

Standardization is the method that is used for rescaling data attributes. The attributes are likely to have a mean value of 0 and a value of the standard deviation of 1. The main objective of standardization is to prompt the mean and standard deviation for the attributes.

50. How to Implement the KNN Classification Algorithm?

Iris dataset is used for implementing the KNN classification algorithm.

```
# KNN classification algorithm
from sklearn.datasets import load_iris
from sklearn.neighbors import KNeighborsClassifier
import numpy as np
from sklearn.model_selection import train_test_split
iris_dataset=load_iris()
A_train, A_test, B_train, B_test = ztrain_test_split(iris_dataset["data"],
iris_dataset["target"], random_state=0)
kn = KNeighborsClassifier(n_neighbors=1)
kn.fit(A_train, B_train)
A_new = np.array([[8, 2.5, 1, 1.2]])
prediction = kn.predict(A_new)
print("Predicted target value: {}\n".format(prediction))
print("Predicted feature name: {}\n".format
(iris_dataset["target_names"][prediction]))
print("Test score: {:.2f}".format(kn.score(A_test, B_test)))
Output:
Predicted Target Name: [0]
Predicted Feature Name: [‘ Setosa’]
Test Score: 0.92
```

The data preprossing techniques:

- **Mean removal** - It involves removing the mean from each feature so that it is centred on zero. Mean removal helps in removing any bias from the features.
- **Feature scaling** - The values of every feature in a data point can vary between random values. So, it is important to scale them so that this matches specified rules.
- **Normalization** - Normalization involves adjusting the values in the feature vector so as to measure them on a common scale. Here, the values of a feature vector are adjusted so that they sum up to 1.
- **Binarization** - Binarization is used to convert a numerical feature vector into a Boolean vector.

- **Normalization**
- `import numpy as np`
- `from sklearn import preprocessing`
- `X_normalized = preprocessing.normalize(X, norm = 'l2')`

We can visualize the data using 2 types of plots :

1. Univariate plots for each individual variable such as Box plot, histogram
2. Multivariate plots such as Scatterplot matrix to understand structured relationship/interactions b/w the variables.

2. What are brute force algorithms? Provide an example.

Brute force algorithms try all possibilities to find a solution. For example, if you were trying to solve a 3-digit pin code, brute force would require you to test all possible combinations from 000 to 999.

One common brute force algorithm is linear search, which traverses an array to check for a match. One disadvantage of brute force algorithms is that they can be inefficient and it's usually more difficult to improve the performance of the algorithm within the framework.

4. What are some ways to handle missing data in Python?

There are two common strategies. **Omission and Imputation**. Omission refers to removing rows or columns with missing values, while imputation refers to adding values to fill in missing observations.

5. What is regression? How would you implement regression in Python?

Regression is a supervised machine learning technique, and it's primarily used to find correlations between variables, as well as make predictions for the dependent variable. Regression algorithms are generally used for predictions, building forecasts, time-series models, or identifying causation.

6. How do you split training and testing datasets in Python?

In Python, you can do this with the Scikit-learn module, using the train_test_split function. This is used to split arrays or matrices into random training and testing datasets. Generally, about 75% of the data will go to the training dataset; however you will likely test different iterations.

```
X_train, X_test, y_train, y_test = train_test_split(data.data, data.target, test_size=0.4)
```

10. Write a function to impute the median price of the selected California cheeses in place of the missing values.

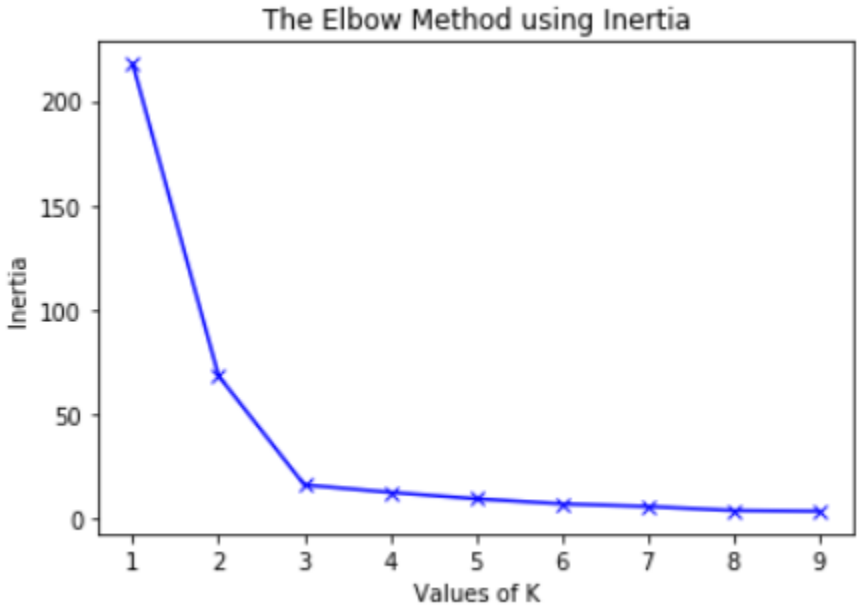
This question requires you to use two built-in pandas methods:

dataframe.column.median()
This method returns the median of a column in a dataframe.

dataframe.column.fillna(value)
This method applies value to all nan values in a given column.

3: How to choose an optimal number of clusters?

By using the Elbow method we decide an optimal number of clusters that our clustering algorithm must try to form. The main principle behind this method is that if we will increase the number of clusters the error value will decrease. But after an optimal number of features, the decrease in the error value is insignificant so, at the point after which this starts to happen, we choose that point as the optimal number of clusters that the algorithm will try to form.



ELBOW METHOD

The optimal number of clusters from the above figure is 3.

6: How do measure the effectiveness of the clusters?

There are metrics like Inertia or Sum of Squared Errors (SSE), Silhouette Score, l1, and l2 scores. Out of all of these metrics, the Inertia or Sum of Squared Errors (SSE) and Silhouette score is a common metrics for measuring the effectiveness of the clusters. Although this method is quite expensive in terms of computation cost. The score is high if the clusters formed are dense and well separated.

7: Why do we take smaller values of the learning rate?

Smaller values of learning rate help the training process to converge more slowly and gradually toward the global optimum instead of fluctuating around it. This is because a smaller learning rate results in smaller updates to the model weights at each iteration, which can help to ensure that the updates are more precise and stable.
If the learning rate is too large, the model weights can update too quickly, which can cause the training process to overshoot the global optimum and miss it entirely.

So, to avoid this oscillation of the error value and achieve the best weights for the model this is necessary to use smaller values of the learning rate.

10: Why do we perform normalization?

To achieve stable and fast training of the model we use normalization techniques to bring all the features to a certain scale or range of values. If we do not perform normalization then there are chances that the gradient will not converge to the global or local minima and end up oscillating back and forth. Read more about it here.

12: What is the difference between upsampling and downsampling?

In the upsampling method, we increase the number of samples in the minority class by randomly selecting some points from the minority class and adding them to the dataset repeat this process till the dataset gets balanced for each class. But here is a disadvantage the training accuracy becomes high as in each epoch model trained more than once in each epoch but the same high accuracy is not observed in the validation accuracy.

In the case of downsampling, we decrease the number of samples in the majority class by selecting some random number of points that are equal to the number of data points in the minority class so that the distribution becomes balanced. In this case, we have to suffer from data loss which may lead to the loss of some critical information as well.

15: What are some of the hyperparameters of the random forest regressor which help to avoid overfitting?

The most important hyper-parameters of a Random Forest are:

- **max_depth** – Sometimes the larger depth of the tree can create overfitting. To overcome it, the depth should be limited.
- **n-estimator** – It is the number of decision trees we want in our forest.
- **min_sample_split** – It is the minimum number of samples an internal node must hold in order to split into further nodes.
- **max_leaf_nodes** – It helps the model to control the splitting of the nodes and in turn, the depth of the model is also restricted.

19: What is one-shot learning?

One-shot learning is a concept in machine learning where the model is trained to recognize the patterns in datasets from a single example instead of training on large datasets. This is useful when we haven't large datasets. It is applied to find the similarity and dissimilarities between the two images.

27: What is the difference between stochastic gradient descent (SGD) and gradient descent (GD)?

In the gradient descent algorithm train our model on the whole dataset at once. But in Stochastic Gradient Descent, the model is trained by using a mini-batch of training data at once. If we are using SGD then one cannot expect the training error to go down smoothly. The training error oscillates but after some training steps, we can say that the training error has gone down. Also, the minima achieved by using GD may vary from that achieved using the SGD. It is observed that the minima achieved by using SGD are close to GD but not the same.

1: Explain the working principle of SVM.

A data set that is not separable in different classes in one plane may be separable in another plane. This is exactly the idea behind the SVM in this a low dimensional data is mapped to high dimensional data so, that it becomes separable in the different classes. A hyperplane is determined after mapping the data into a higher dimension which can separate the data into categories. SVM model can even learn non-linear boundaries with the objective that there should be as much margin as possible between the categories in which the data has been categorized. To perform this mapping different types of kernels are used like radial basis kernel, gaussian kernel, polynomial kernel, and many others.

3: Explain some measures of similarity which are generally used in Machine learning.

Some of the most commonly used similarity measures are as follows:

- **Cosine Similarity** – By considering the two vectors in n – dimension we evaluate the cosine of the angle between the two. The range of this similarity measure varies from [-1, 1] where the value 1 represents that the two vectors are highly similar and -1 represents that the two vectors are completely different from each other.
- **Euclidean or Manhattan Distance** – These two values represent the distances between the two points in an n-dimensional plane. The only difference between the two is in the way the two are calculated.
- **Jaccard Similarity** – It is also known as IoU or Intersection over union it is widely used in the field of object detection to evaluate the overlap between the predicted bounding box and the ground truth bounding box.

6: What is the difference between L1 and L2 regularization? What is their significance?

L1 regularization: In L1 regularization also known as Lasso regularization in which we add the sum of absolute values of the weights of the model in the loss function.
In L1 regularization weights for those features which are not at all important are penalized to zero so, in turn, we obtain feature selection by using the L1 regularization technique.

L2 regularization: In L2 regularization also known as Ridge regularization in which we add the square of the weights to the loss function. In both of these regularization methods, weights are penalized but there is a subtle difference between the objective they help to achieve.
In L2 regularization the weights are not penalized to 0 but they are near zero for irrelevant features. It is often used to prevent overfitting by shrinking the weights towards zero, especially when there are many features and the data is noisy.

9: Does the accuracy score always a good metric to measure the performance of a classification model?

No, there are times when we train our model on an imbalanced dataset the accuracy score is not a good metric to measure the performance of the model. In such cases, we use precision and recall to measure the performance of a classification model. Also, f1-score is another metric that can be used to measure performance but in the end, f1-score is also calculated using precision and recall as the f1-score is nothing but the harmonic mean of the precision and recall.

15: What is Linear Discriminant Analysis?

LDA is a supervised machine learning dimensionality reduction technique because it uses target variables also for dimensionality reduction. It is commonly used for classification problems. The LDA mainly works on two objectives:

- Maximize the distance between the means of the two classes.
- Minimize the variation within each class.

18: Whether the metric MAE or MSE or RMSE is more robust to the outliers.

Out of the above three metrics, MAE is robust to the outliers as compared to the MSE or RMSE. The main reason behind this is because of Squaring the error values. In the case of an outlier, the error value is already high and then we squared it which results in an explosion in the error values more than expected and creates misleading results for the gradient.

19: Why removing highly correlated features are considered a good practice?

When two features are highly correlated, they may provide similar information to the model, which may cause overfitting. If there are highly correlated features in the dataset then they unnecessarily increase the dimensionality of the feature space and sometimes create the problem of the curse of dimensionality. If the dimensionality of the feature space is high then the model training may take more time than expected, it will increase the complexity of the model and chances of error. This somehow also helps us to achieve data compression as the features have been removed without much loss of data.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression
```

NumPy

- `np.array()`: Create an array from a list or tuple.
- `np.zeros()`: Create an array of zeros.
- `np.ones()`: Create an array of ones.
- `np.arange()`: Create an array with a range of values.
- `array.shape`: Get the shape of the array.
- `array.reshape()`: Reshape the array.
- `array.flatten()`: create a 1-dimensional array (vector) from a multi-dimensional array
- `np.concatenate()`: Concatenate arrays.
- `np.add()`, `np.subtract()`, `np.multiply()`, `np.divide
- `np.dot()`, `np.matmul()`: Perform matrix multiplication.
- `np.sum()`, `np.mean()`, `np.std()`: Calculate sum, mean, and standard deviation.
- `np.min()`, `np.max()`: Find minimum and maximum values.
- `np.sin()`, `np.cos()`, `np.exp()`, `np.log()`: Apply functions element-wise.
- `np.sqrt()`, `np.power()`: Calculate square root and power.
- `np.equal()`, `np.not_equal()`, `np.greater()`, `np.less()`:
- `np.logical_and()`, `np.logical_or()`, `np.logical_not()`:
- `array.T`: Transpose the array.
- `np.transpose()`: Alternate method for array transposition.
- `np.random.rand()`: Generate random values from a uniform distribution.
- `np.random.randn()`: Generate random values from a standard normal distribution.
- `np.random.randint()`: Generate random integers.
- `np.random.choice()`: Randomly sample elements from an array.
- `np.linalg.inv()`: Calculate the inverse of a matrix.
- `np.linalg.eig()`: Calculate eigenvalues and eigenvectors.
- `np.linalg.solve()`: Solve linear systems of equations.

Pandas

- `pd.DataFrame()`: Create a DataFrame from data, lists, or dictionaries.

The `pd.DataFrame()` function in pandas is used to create a DataFrame, which is a 2-dimensional labeled data structure similar to a table in a database or a spreadsheet. It can hold data of various types and is a powerful tool for data manipulation and analysis. Here's an illustration of how to use `pd.DataFrame()`:

```
import pandas as pd

data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Age': [25, 30, 28],
        'City': ['New York', 'San Francisco', 'Los Angeles']}
df = pd.DataFrame(data)
```

- `pd.read_csv()`, `pd.read_excel()`: Load data from CSV files into a DataFrame.
- `pd.read_sql()`: Load data from a SQL database into a DataFrame.
- `df.head()`, `df.tail()`: Display the first or last few rows of the DataFrame.
- `df.info()`: Display information about the DataFrame's data types and non-null
- `df.describe()`: Generate summary statistics for numerical columns.
- `df[column]`: Access a column by its label.
- `df.loc[row_label, column_label]`: Access a value using labels.
- `df.iloc[row_index, column_index]`: Access a value using integer indices.
- `df.iloc[:, column_index]`: Select all rows for a specific column.
- `df.drop()`: Remove rows or columns from the DataFrame.
- `df.fillna()`: Fill missing values in the DataFrame.
- `df.replace()`: Replace specific values in the DataFrame.
- `pd.merge()`: Merge two DataFrames based on common columns.
- `df.join()`: Join two DataFrames using their indices.

- `df.sort_values()`: Sort the DataFrame by one or more columns.
- `df.to_csv()`, `df.to_excel()`: Save the DataFrame to CSV or Excel format.
- `df.to_sql()`: Save the DataFrame to a SQL database.
- `df.apply()`: Apply a function element-wise or along an axis.
- `pd.to_datetime()`: Convert a column to a datetime format.
- `df[df['column'] > value]`: Filter rows based on a condition.

```
import pandas as pd
# Create a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie', 'David'],
        'Age': [25, 30, 28, 22],
        'City': ['New York', 'San Francisco', 'Los Angeles', 'Chicago']}
df = pd.DataFrame(data)
# Filter rows where Age is greater than or equal to 28
filtered_df = df[df['Age'] >= 28]
```

- `df.query('condition')`: Perform SQL-like queries on the DataFrame.

```
import pandas as pd
# Create a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie', 'David'],
        'Age': [25, 30, 28, 22],
        'City': ['New York', 'San Francisco', 'Los Angeles', 'Chicago']}
df = pd.DataFrame(data)
# Query the DataFrame to filter rows where Age is greater than or equal to 28
filtered_df = df.query('Age >= 28')
```

- `df.groupby()`: Group data based on a column's values.

```
import pandas as pd

# Create a DataFrame

data = {'Name': ['Alice', 'Bob', 'Alice', 'Bob', 'Charlie'],

        'Age': [25, 30, 28, 32, 28],

        'City': ['New York', 'San Francisco', 'New York', 'San Francisco', 'Los Angeles']}

df = pd.DataFrame(data)

# Grouping by 'City' column and calculating mean age within each city

grouped = df.groupby('City')['Age'].mean()
```

matplotlib.pyplot

- `plt.plot()`: Create a line plot.
- `plt.scatter()`: Create a scatter plot.
- `plt.plot_date()`: Create a line plot with dates on the x-axis.
- `plt.bar()`: Create vertical bar plots.
- `plt.hist()`: Create histograms.
- `plt.pie()`: Create pie charts.
- `plt.xlabel()`, `plt.ylabel()`: Add labels to x and y axes.
- `plt.title()`: Add a title to the plot.
- `plt.legend()`: Add legends to the plot.
- `plt.text()`: Add text to the plot.
- `plt.savefig()`: Save the plot to a file.
- `plt.show()`: Display the plot on the screen.