Section A

1. Define Machine learning with its objectives

Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed.

The primary purpose of machine learning is to discover patterns in the user data and then make predictions based on these and intricate patterns for answering business questions and solving business problems.

2. Define goal of the support vector machine (SVM).

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen.

3. List out different algorithms can be classified under Association Rule Learning Algorithms?

Association rule learning is a type of unsupervised learning technique that checks for the dependency of one data item on another data item and maps accordingly so that it can be more profitable. It tries to find some interesting relations or associations among the variables of dataset. It is based on different rules to discover the interesting relations between variables in the database.

The association rule learning is one of the very important concepts of machine learning, and it is employed in **Market Basket analysis, Web usage mining, continuous production, etc.**

Association rule learning can be divided into three types of algorithms: 1. Apriori 2. Eclat 3. F-P Growth Algorithm

4. **Define any algorithm you know in to solve a problem in Reinforcement Learning.**

There are two important learning models in reinforcement learning:

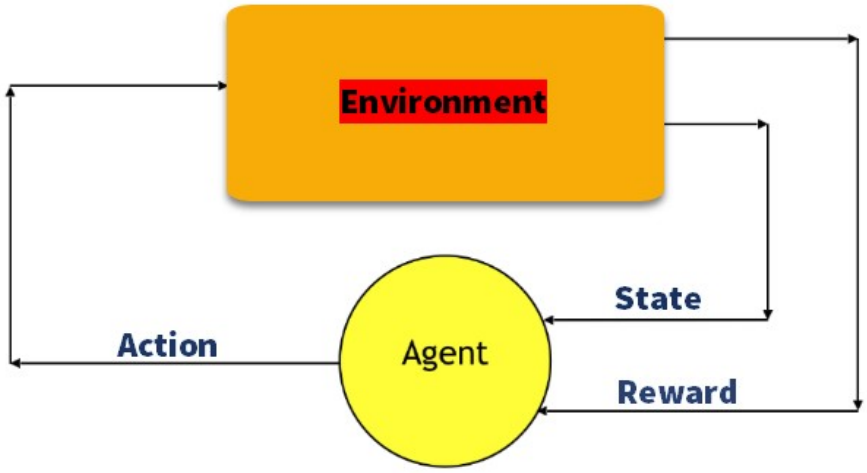
- Markov Decision Process
- Q learning

Markov Decision Process

The following parameters are used to get a solution:

- Set of actions- A
- Set of states -S
- Reward- R
- Policy- n
- Value- V

The mathematical approach for mapping a solution in reinforcement Learning is recon as a Markov Decision Process or (MDP).



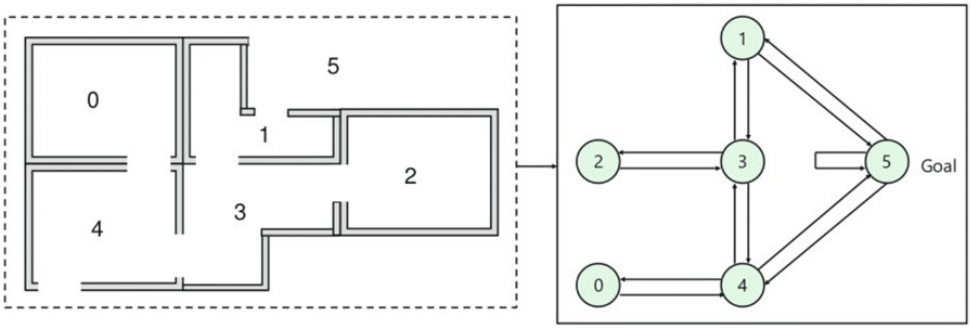
Q-Learning

Q learning is a value-based method of supplying information to inform which action an agent should take.

Let's understand this method by the following example:

- There are five rooms in a building which are connected by doors.

- Each room is numbered 0 to 4
- The outside of the building can be one big outside area (5)
- Doors number 1 and 4 lead into the building from room 5



Next, you need to associate a reward value to each door:

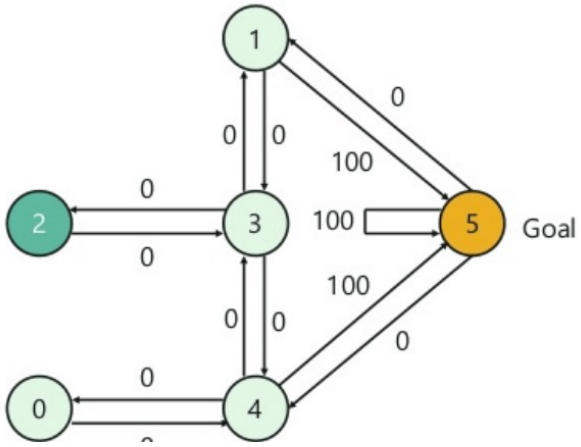
- Doors which lead directly to the goal have a reward of 100
- Doors which is not directly connected to the target room gives zero reward
- As doors are two-way, and two arrows are assigned for each room
- Every arrow in the above image contains an instant reward value

Explanation:

In this image, you can view that room represents a state

Agent's movement from one room to another represents an action

In the below-given image, a state is described as a node, while the arrows show the action.



For example, an agent traverse from room number 2 to 5

- Initial state = state 2
- State 2-> state 3
- State 3 -> state (2,1,4)
- State 4-> state (0,5,3)
- State 1-> state (5,3)
- State 0-> state 4

5. **Define Agglomerative Clustering and divisive clustering.**

   Agglomerative Clustering: Also known as bottom-up approach or hierarchical agglomerative clustering (HAC). A structure that is more informative than the unstructured set of clusters returned by flat clustering. This clustering algorithm does not require us to prespecify the number of clusters. Bottom-up algorithms treat each data as a singleton cluster at the outset and then successively agglomerates pairs of clusters until all clusters have been merged into a single cluster that contains all data.

   Divisive clustering: Also known as a top-down approach. This algorithm also does not require to prespecify the number of clusters. Top-down clustering requires a method for splitting a cluster that contains the whole data and proceeds by splitting clusters recursively until individual data have been split into singleton clusters.

Section B

6. Differentiate between Supervised, Unsupervised and Reinforcement Learning

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

7. How would you detect overfitting in Linear Models?

Overfitting is a modeling error that occurs when a function or model is too closely fit the training set and getting a drastic difference of fitting in test set. Overfitting the model generally takes the form of making an overly complex model to explain Model behavior in the data under study.

So the first step to finding the Overfitting is to split the data into the Training and Testing set. If our model does much better on the training set than on the test set, then we're likely overfitting. The performance can be measured using the percentage of accuracy observed in both data sets to conclude on the presence of overfitting. If the model performs better on the training set than on the test set, it means that the model is likely overfitting.

In linear regression overfitting occurs when the model is "too complex". This usually happens when there are a large number of parameters compared to the number of observations.

8. **Elaborate relationship between k-Means Clustering and PCA?**

Principal Component Analysis (PCA) is a tool for dimension reduction. This technique is to transform the larger dataset into a smaller dataset by identifying the correlations and patterns with preserving most of the valuable information.

This is need for feature selection of a model. PCA aims to capture valuable information explaining high variance which results in providing the best accuracy.

K-Means Clustering:

It is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups(clusters) where each data point belongs to only one group.

K-means clustering uses “centroids”, K different randomly-initiated points in the data, and assigns every data point to the nearest centroid. After every point has been assigned, the centroid is moved to the average of all of the points assigned to it.

Relationship between PCA and K Means Clustering:

Both uses are in dimensionality reduction for visualizing patterns in data from parameters(variables).PCA in conjunction with K-means is a powerful method for visualizing high dimensional data.

k-means tries to find the least-squares partition of the data. PCA finds the least-squares cluster membership vector. The first Eigenvector has the largest variance, therefore splitting on this vector (which resembles cluster membership, not input data coordinates!) means maximizing between cluster variance.

However, when we employ PCA prior to using K-means we can visually separate almost the entire data set. That was one of the biggest goals of PCA - to reduce the number of variables by combining them into bigger, more meaningful features. Not only that, but they are 'orthogonal' to each other.

9. Compare Reinforcement Learning and Supervised Learning.

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

10.

10	The values of independent variable x and dependent value y are given below: X: 1, 3, 5, 7, 8 Y: 3, 4, 2, 5, 7 Find the regression line y=ax+b. Estimate the value of y when x is 11.

Concept:

The normal equation for Fitting a straight line by the **least square method** is:

$\Sigma y = na + b \Sigma x$

$\Sigma xy = a \Sigma x + b \Sigma x^2$

Where

n = Total number of observations, a and b are the coefficients.

By solving the above two equations coefficients a and b can be obtained.

Given Data and Calculation:

x	y	x^2	xy
5	16	25	80
2	10	4	20
4	13	16	52
3	12	9	36
Σx = 14	Σy = 51	Σx^2 = 54	Σxy = 188

n = 4 So

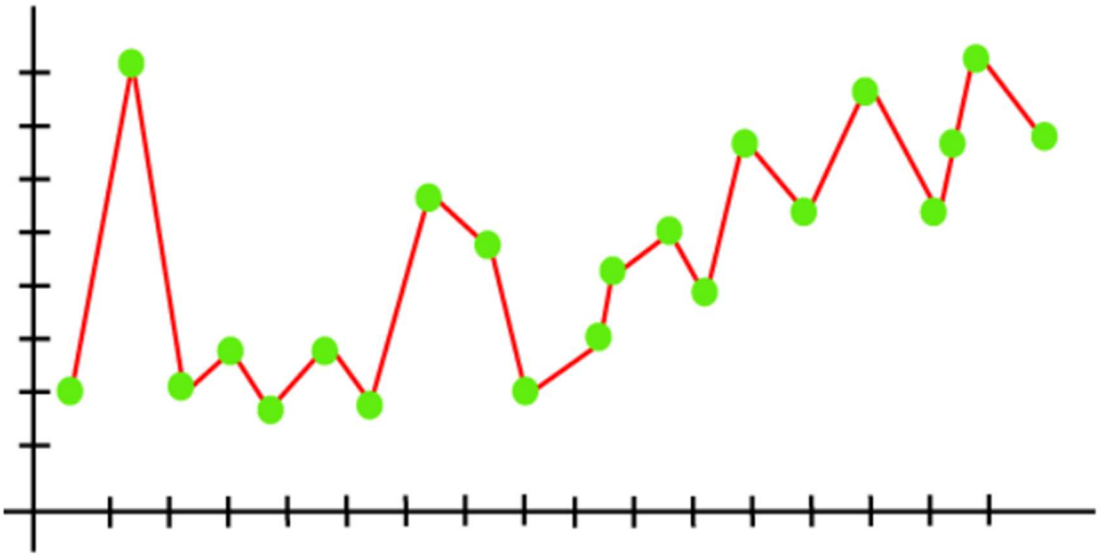
51 = 4a + 14b

188 = 14a + 54b

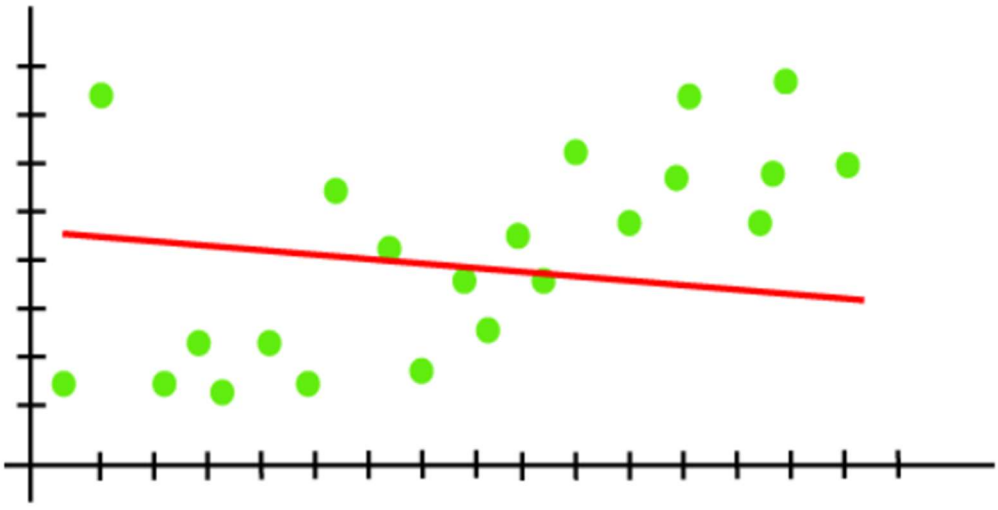
Solving the above two equations **a = 6.1 and b = 1.9**

11. Compare overfitting and underfitting with an intuitive explanation of the Bias-Variance Tradeoff.

Overfitting is a modeling error that occurs when a function or model is too closely fit the training set and getting a drastic difference of fitting in test set. Overfitting the model generally takes the form of making an overly complex model to explain Model behavior in the data under study.



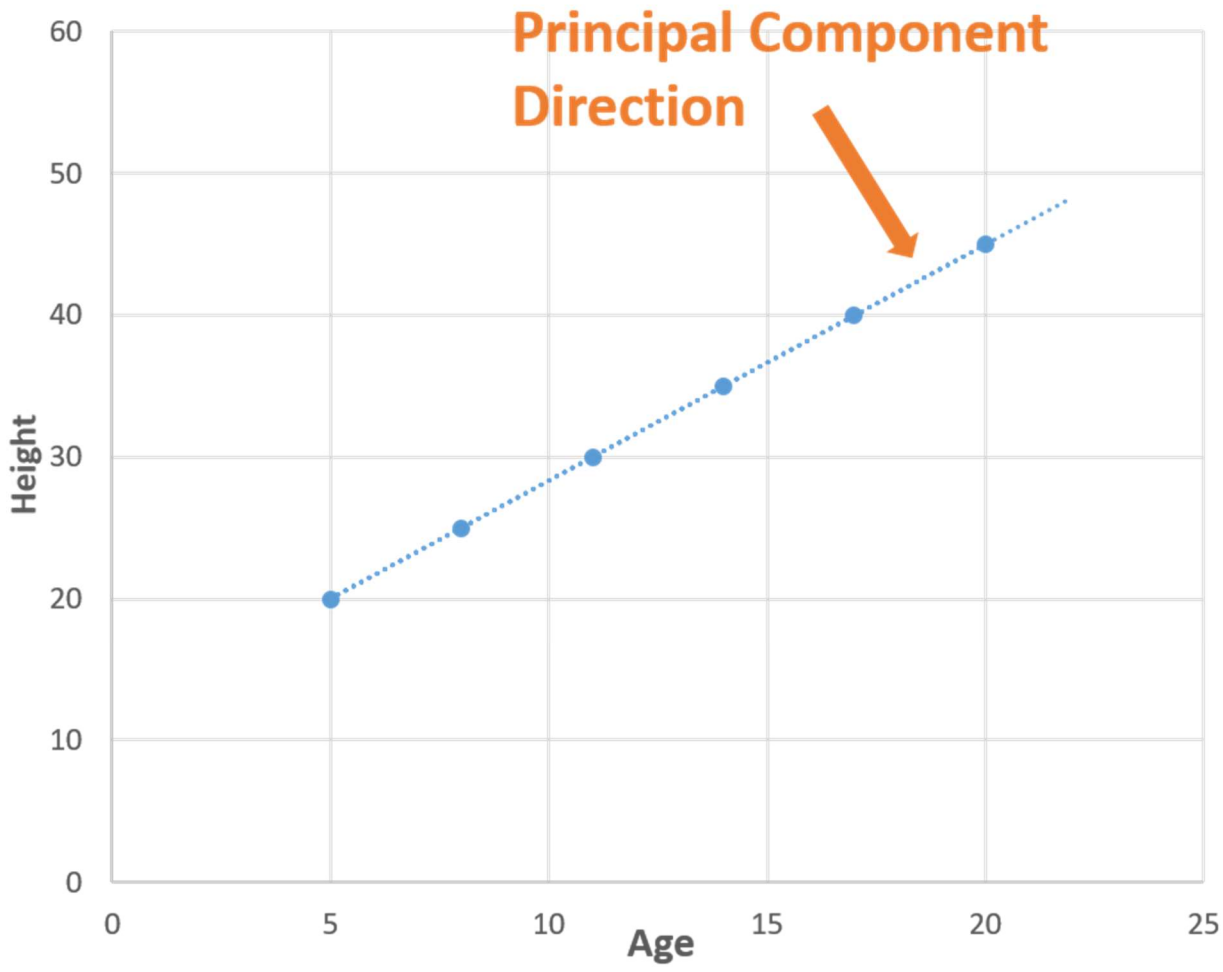
A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data, i.e., it only performs well on training data but performs poorly on testing data.



12. **Scrutinize that Principal Component Analysis (PCA) is used for Dimensionality Reduction with an example.**

Principal Component Analysis (PCA) is a tool for dimension reduction. This technique is to transform the larger dataset into a smaller dataset by identifying the correlations and patterns with preserving most of the valuable information.

This is need for feature selection of a model. PCA aims to capture valuable information explaining high variance which results in providing the best accuracy.



Now consider a line (blue dotted line) that is passing through all the points. The blue dotted line is capturing all the information so, we can replace 'Age' and 'Weight' (2 Dimensions) with the blue dotted line (1 Dimension) without losing

any information, and in this way, we have done dimensionality reduction (2 dimensions to 1 dimension). The blue dotted line is called **Principal Component.**

Steps for PCA algorithm

1. **Getting the dataset**

   Firstly, we need to take the input dataset and divide it into two subparts X and Y, where X is the training set, and Y is the validation set.
2. **Representing data into a structure**

   Now we will represent our dataset into a structure. Such as we will represent the two-dimensional matrix of independent variable X. Here each row corresponds to the data items, and the column corresponds to the Features. The number of columns is the dimensions of the dataset.
3. **Standardizing the data**

   In this step, we will standardize our dataset. Such as in a particular column, the features with high variance are more important compared to the features with lower variance.

   If the importance of features is independent of the variance of the feature, then we will divide each data item in a column with the standard deviation of the column. Here we will name the matrix as Z.
4. **Calculating the Covariance of Z**

   To calculate the covariance of Z, we will take the matrix Z, and will transpose it. After transpose, we will multiply it by Z. The output matrix will be the Covariance matrix of Z.
5. **Calculating the Eigen Values and Eigen Vectors**

   Now we need to calculate the eigenvalues and eigenvectors for the resultant covariance matrix Z. Eigenvectors or the covariance matrix are the

directions of the axes with high information. And the coefficients of these eigenvectors are defined as the eigenvalues.

6. **Sorting the Eigen Vectors**

   In this step, we will take all the eigenvalues and will sort them in decreasing order, which means from largest to smallest. And simultaneously sort the eigenvectors accordingly in matrix P of eigenvalues. The resultant matrix will be named as P*.

7. **Calculating the new features Or Principal Components**

   Here we will calculate the new features. To do this, we will multiply the P* matrix to the Z. In the resultant matrix Z*, each observation is the linear combination of original features. Each column of the Z* matrix is independent of each other.

8. **Remove less or unimportant features from the new dataset.**

   The new feature set has occurred, so we will decide here what to keep and what to remove. It means, we will only keep the relevant or important features in the new dataset, and unimportant features will be removed out.