**FSDS MAY BATCH 2022(ML Assignment -2)**

**Submitted by: Shubham Tiwari**

Q1:What is the concept of human learning? Please give two examples.

Ans:Human learning is basically an approach which helps us to understand about the complex environment and enables us to work effectively in that environment .It is basically purposive and goal directive. The two examples can be driving a motor car, memorizing poems etc.

Q2: What different forms of human learning are there? Are there any machine learning equivalents?

Ans: There are basically 3 different forms:

1)      Learning through association  - Classical Conditioning

2)      Learning through consequences – Operant Conditioning

3)      Learning through observation – Modeling/Observational Learning

Yes, the machine learning equivalent can be supervised learning, unsupervised learning and reinforcement learning.

Q3: What is machine learning, and how does it work? What are the key responsibilities of machine learning?

Ans: Machine Learning basically provides statistical tool to analyze, visualize, perform predictions and other many task with the help of data. It basically helps to solve real world problems, unlike the hard coding machine learning algorithm learn from the data.It was introduced around 1950’s by Alan Turing. Machine learning uses two main techniques: Supervised and unsupervised learning techniques.

The responsibilities of machine learning can be :

* To create and construct methods and plans for machine learning.
* Performing Statisticals analysis and improve model.
* To improve and broaden current ML frameworks and libraries.
* To investigate, test, and put into practice appropriate ML tools and algorithms.
* To evaluate the application cases and problem-solving potential of ML algorithms and rank them according to success rates.

Q4: Define the terms “penalty” and “reward” in the context of reinforcement learning.

Ans: **Reward:** For every action made, the agent receives a reward from the environment. A reward could be positive or negative, depending on the action.

**Penalty:** This type of machine learning uses a reward-penalty method to teach an AI system. If it makes the right move, it gets rewarded. If it makes a mistake, it receives a penalty.

Q5: Explain the term “learning as a search”?

Ans: It can be defined as “Learning can be viewed as a search through the space of all sentences in a concept description language for a sentence that best describes the data. Alternatively, it can be viewed as a search through all hypotheses in a hypothesis space. In either case, a generality relation usually determines the structure of the search space.”

Q6: What are the various goals of machine learning? What is the relationship between these and human learning?

Ans: The various goals of machine learning are:

1) To make the computers smarter, more intelligent. The more direct objective in this aspect is to develop systems (programs) for specific practical learning tasks in application domains.

2) To develop computational models of human learning process and perform computer simulations. The study in this aspect is also called cognitive modeling.

3) To explore new learning methods and develop general learning algorithms independent of applications.

We compare the human learning with machine learning along the dimensions of speed, ability to transfer, and others. which shows that machine learning is both an opportunity and challenge, in the sense that we can hope to discover ways for machine to learn which are better than ways human learn (the opportunity), and that there are amply amount of difficulties to be overcome in order to make machines learn (the challenge).

**Dimension Human Learning Machine Learning**

1)Speed Slow Slow-hope to find tricks for machine to learn fast

2)Ability to transfer No copy mechanism Easy to copy

3)Require repetition Yes Yes/No

4)Error-prone Yes Yes

5)Noise-tolerant Yes No

Q7: Illustrate the various elements of machine learning using a real-life illustration.

Ans: The various elements of machine learning using a real-life illustration are:

**1)** **Image recognition**

Image recognition is a well-known and widespread example of machine learning in the real world. It can identify an object as a digital image, based on the intensity of the pixels in black and white images or colour images.

Real-world examples of image recognition:

* Label an x-ray as cancerous or not
* Assign a name to a photographed face (aka “tagging” on social media)
* Recognise handwriting by segmenting a single letter into smaller images.

**2) Speech recognition**

Machine learning can translate speech into text. Certain software applications can convert live voice and recorded speech into a text file. The speech can be segmented by intensities on time-frequency bands as well.

Real-world examples of speech recognition:

* Voice search
* Voice dialling
* Appliance control

Some of the most common uses of speech recognition software are devices like Google Home or Amazon Alexa.

**3)** **Medical diagnosis**

Machine learning can help with the diagnosis of diseases. Many physicians use chatbots with speech recognition capabilities to discern patterns in symptoms.

Real-world examples for medical diagnosis:

* Assisting in formulating a diagnosis or recommends a treatment option
* Oncology and pathology use machine learning to recognise cancerous tissue
* Analyse bodily fluids

In the case of rare diseases, the joint use of facial recognition software and machine learning helps scan patient photos and identify phenotypes that correlate with rare genetic diseases.

**4) Statistical arbitrage**

Arbitrage is an automated trading strategy that’s used in finance to manage a large volume of securities. The strategy uses a trading algorithm to analyse a set of securities using economic variables and correlations.

Real-world examples of statistical arbitrage:

* Algorithmic trading which analyses a market microstructure
* Analyse large data sets
* Identify real-time arbitrage opportunities

Machine learning optimises the arbitrage strategy to enhance results.

**5) Predictive analytics**

Machine learning can classify available data into groups, which are then defined by rules set by analysts. When the classification is complete, the analysts can calculate the probability of a fault.

Real-world examples of predictive analytics:

* Predicting whether a transaction is fraudulent or legitimate
* Improve [prediction systems](https://www.salesforce.com/eu/products/marketing-cloud/platform/predictive-internet-intelligence) to calculate the possibility of fault

Predictive analytics is one of the most promising examples of machine learning. It's applicable for everything; from product development to real estate pricing.

Q8: Provide an example of the abstraction method.

Ans: In abstraction, the users are familiar with the purpose of the class's methods, but they don't know how they solve the purpose, which means that they know the inputs and expected outputs, but the inner working is hidden.

There are plenty of real-life examples, like when we press a button on a TV remote to change the channel; we don’t know how it does that. We are just interested in the fact that when we press a particular button, it changes the channel.

We can use the following syntax to create an abstract class in Python:

from abc import ABC

class <Abstract\_Class\_Name>(ABC):

# body of the class

Q9: What is the concept of generalization? What function does it play in the machine learning process?

Ans: In machine learning, generalization is a definition to demonstrate how well is a trained model to classify or forecast unseen data. Training a generalized machine learning model means, in general, it works for all subset of unseen data. An example is when we train a model to classify between dogs and cats. If the model is provided with dogs images dataset with only two breeds, it may obtain a good performance. But, it possibly gets a low classification score when it is tested by other breeds of dogs as well. This issue can result to classify an actual dog image as a cat from the unseen dataset. Therefore, data diversity is very important factor in order to make a good prediction. In the sample above, the model may obtain 85% performance score when it is tested by only two dog breeds and gains 70% if trained by all breeds. However, the first possibly gets a very low score (e.g. 45%) if it is evaluated by an unseen dataset with all breed dogs. This for the latter can be unchanged given than it has been trained by high data diversity including all possible breeds.

Q10: What is classification, exactly? What are the main distinctions between classification and regression?

Ans: Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.

The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y).

**Example:** The best example to understand the Classification problem is Email Spam Detection. The model is trained on the basis of millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

The difference between them can be understood by the following table:

|  |  |
| --- | --- |
| **Regression Algorithm** | **Classification Algorithm** |
| In Regression, the output variable must be of continuous nature or real value. | In Classification, the output variable must be a discrete value. |
| The task of the regression algorithm is to map the input value (x) with the continuous output variable(y). | The task of the classification algorithm is to map the input value(x) with the discrete output variable(y). |
| Regression Algorithms are used with continuous data. | Classification Algorithms are used with discrete data. |
| In Regression, we try to find the best fit line, which can predict the output more accurately. | In Classification, we try to find the decision boundary, which can divide the dataset into different classes. |
| Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc. | Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc. |
| The regression Algorithm can be further divided into Linear and Non-linear Regression. | The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier. |

Q11: What is regression, and how does it work? Give an example of a real-world problem that was solved using regression.

Ans: For clear understanding of regression following points must be noted:

* A regression is a statistical technique that relates a dependent variable to one or more independent (explanatory) variables.
* A regression model is able to show whether changes observed in the dependent variable are associated with changes in one or more of the explanatory variables.
* It does this by essentially fitting a best-fit line and seeing how the data is dispersed around this line.
* Regression captures the correlation between variables observed in a data set, and quantifies whether those correlations are statistically significant or not.

**Example:** 1) Linear Regression can be also used to assess risk in financial services or insurance domain. For example, a car insurance company might conduct a linear regression to come up with a suggested premium table using predicted claims to Insured Declared Value ratio. The risk can be assessed based on the attributes of the car, driver information or demographics. The results of such an analysis might guide important business decisions.

2)Financial forecasting (like house price estimates, or stock prices)

3)Sales and promotions forecasting

3)Testing automobiles

4)Weather analysis and prediction

5)Time series forecasting.

Q12: Describe the clustering mechanism in detail.

Ans: It is basically a type of [unsupervised learning method](https://www.geeksforgeeks.org/supervised-unsupervised-learning/). An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

It can be defined as "A way of grouping the data points into different clusters, consisting of similar data points. The objects with the possible similarities remain in a group that has less or no similarities with another group." It is an unsupervised learning method, hence no supervision is provided to the algorithm, and it deals with the unlabeled dataset. The clustering technique is commonly used for statistical data analysis.

The clustering technique can be widely used in various tasks. Some most common uses of this technique are:

* Market Segmentation
* Statistical data analysis
* Social network analysis
* Image segmentation
* Anomaly detection, etc.

Apart from these general usages, it is used by the **Amazon** in its recommendation system to provide the recommendations as per the past search of products. **Netflix** also uses this technique to recommend the movies and web-series to its users as per the watch history

* **Types of Clustering Methods:**

The clustering methods are broadly divided into **Hard clustering** (datapoint belongs to only one group) and **Soft Clustering** (data points can belong to another group also). But there are also other various approaches of Clustering exist. Below are the main clustering methods used in Machine learning:

1. **Partitioning Clustering**
2. **Density-Based Clustering**
3. **Distribution Model-Based Clustering**
4. **Hierarchical Clustering**
5. **Fuzzy Clustering**

* **Clustering Algorithms:** The clustering algorithm is based on the kind of data that we are using. Such as, some algorithms need to guess the number of clusters in the given dataset, whereas some are required to find the minimum distance between the observation of the dataset.

**1)K-Means algorithm**: The k-means algorithm is one of the most popular clustering algorithms. It classifies the dataset by dividing the samples into different clusters of equal variances. The number of clusters must be specified in this algorithm. It is fast with fewer computations required, with the linear complexity of O(n).

**2)Mean-shift algorithm**: Mean-shift algorithm tries to find the dense areas in the smooth density of data points. It is an example of a centroid-based model, that works on updating the candidates for centroid to be the center of the points within a given region.

**3)DBSCAN Algorithm**: It stands for Density-Based Spatial Clustering of Applications with Noise. It is an example of a density-based model similar to the mean-shift, but with some remarkable advantages. In this algorithm, the areas of high density are separated by the areas of low density. Because of this, the clusters can be found in any arbitrary shape.

Q13: Make brief observations on **two** of the following topics:

i. Machine learning algorithms are used

Ans: Although there are various machine learning algorithms but some of them are listed :

**1. Linear regression**

It is one of the most popular Supervised Machine Learning algorithms in Python that maintains an observation of continuous features and based on it, predicts an outcome. It establishes a relationship between dependent and independent variables by fitting a best line. This **best fit line is represented by a linear equation Y=aX+b,** commonly called the regression line.

In this equation,

**Y – Dependent variable**

**a- Slope**

**X – Independent variable**

**b- Intercept**

The regression line is the line that **fits best in the equation to supply a relationship between the dependent and independent variables**.

**2. Decision Trees**

A decision tree is built by repeatedly asking questions to the partition data. The aim of the decision tree algorithm is to **increase the predictiveness at each level of partitioning so that the model is always updated with information about the dataset.**

Even though it is a **Supervised Machine Learning algorithm**, it is used mainly for **classification rather than regression**. The good part about this machine learning algorithm is that**it works on both continuous dependent and categorical variables.**

**3. Logistic regression**

A **supervised machine learning algorithm in Python**that is used in estimating discrete values in binary, e.g: 0/1, yes/no, true/false. This is based on a set of independent variables. This algorithm is used to **predict the probability of an event’s occurrence by fitting that data into a logistic curve or logistic function.**This is why it is also called logistic regression.

Logistic regression, also called as Sigmoid function, takes in any real valued number and then maps it to a value that falls between 0 and 1.

Sigmoid Function is defined as,

**f(x) = L / 1+e^(-x)**

x: domain of real numbers

L: curve’s max value

**4. Support Vector Machines (SVM)**

This is one of the most important machine learning algorithms in Python which is mainly used for **classification but can also be used for regression tasks**. In this algorithm, each data item is plotted as a point in n-dimensional space, where **n denotes the number of features you have, with the value of each feature as the value of a particular coordinate.**

SVM does the **distinction of these classes by a decision boundary.**

**5. Naive Bayes**

Naive Bayes is a **supervised machine learning algorithm used for classification tasks**. This is one of the reasons it is also called a Naive Bayes Classifier. It assumes that features are independent of one another and there exists no correlation between them. But as these assumptions hold no truth in real life, this algorithm is called ‘naive’.

This algorithm works on Bayes’ theorem which is:

**p(A|B) = p(A) . p(B|A) / p(B)**

In this,

p(A): Probability of event A

p(B): Probability of event B

p(A|B): Probability of event A given event B has already occurred

p(B|A): Probability of event B given event A has already occurred

ii. Studying under supervision

Ans:xxxxxxxxxxxx

iii. Studying without supervision

Ans:xxxxxxxxxxxx

iv. Reinforcement learning is a form of learning based on positive reinforcement.

Ans: Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.

* **Main points** in Reinforcement learning :
* **Input:** The input should be an initial state from which the model will start
* **Output**: There are many possible outputs as there are a variety of solutions to a particular problem
* **Training**: The training is based upon the input, The model will return a state and the user will decide to reward or punish the model based on its output.
* The model keeps continues to learn.
* The best solution is decided based on the maximum reward.

**Types of Reinforcement:** There are two types of Reinforcement: 

1. **Positive**   
   Positive Reinforcement is defined as when an event, occurs due to a particular behavior, increases the strength and the frequency of the behavior. In other words, it has a positive effect on behavior.

* **Advantages** of reinforcement learning are:
  + Maximizes Performance
  + Sustain Change for a long period of time
  + Too much Reinforcement can lead to an overload of states which can diminish the results

**2.Negative –**   
Negative Reinforcement is defined as strengthening of behavior because a negative condition is stopped or avoided.

* **Advantages** of reinforcement learning:
  + Increases Behavior
  + Provide defiance to a minimum standard of performance
  + It Only provides enough to meet up the minimum behavior