**FSDS MAY BATCH 2022(ML Assignment -1)**

**Submitted by: Shubham Tiwari**

Q1: What does one mean by the term “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.

Q2: Can you think of 4 distinct types of issues where it shines?

Ans:Machine learning is very helpful in solving real worlds problems such as in case of **spam detection in mail,** **various disease disorders in hospitals,** **credit card fault detection i.e transactions** and recently **automatic AI vehicles** came into existence.

Q3: What is a labeled training set, and how does it work?

Ans:A labelled training set is basically used in supervised machine learning algorithms i.e in regression ,SVR, Decision Tree etc and it is a collection of data where one of the features in the dataset indicates the class where it belongs to .

Q4: What are the two most important tasks that are supervised?

Ans:We basically supervised tasks into two types i.e **regression** and **classification**.In regression problem our prediction is a scaler value but in case of classification problem our prediction or output is the form of 0 or1.

Q5: .Can you think of four examples of unsupervised tasks?

Ans: The unsupervised tasks can be such as **clustering algorithms, K means ,dimensionality reduction, data visualization ,DB scan** etc.

Q6: State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?

Ans: The best machine learning algorithm that would be best to make a robot walk through various unfamiliar terrains is **reinforced learning** where the robot can learn from the response of the terrain to optimize itself.

Q7: Which algorithm will you use to divide your customers into different groups?

Ans: The best algorithm if we divide our customers into different groups can be either supervised learning (if the group has a known label)and we use classification method inside it or Unsupervised learning (if we do not know the group label)and inside it we can use clustering method.

Q8:Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?

Ans: In case of the spam detection we already know the label of group so it is a supervised machine learning algorithm.

Q9: What is the concept of an online learning system?

Ans: In this system here we give the data in small bunches or streams and machine learns from it,this technique is also called mini batch approach.

Q10:What is out-of-core learning, and how does it differ from core learning?

Ans: Out of core learning is basically used that can handle data that cannot fit into computer memory ,it uses online learning system to input the data in small bits. In core learning all the concepts come into play i.e it fits all the data precisizely.

Q11:What kind of learning algorithm makes predictions using a similarity measure?

Ans: Instance based learning algorithm makes predictions using a similarity measure,here the algorithm learns the example and then uses the similarity measures to generalize the data.

Q12: What is the difference between a model parameter and a hyperparameter in a learning algorithm?

Ans: Hyperparameter is a parameter of the learning algorithm not the model.The clear difference can be understood by following:

* **Model parameter:**

1) These are essential for

making predictions.

2) These are dependent on the

dataset, which is used for training.

3) The values of parameters can be

estimated by the optimization

algorithms, such as Gradient

Descent.

* **Hyperparameter:**

1) Hyperparameters are essential for optimizing the model.

2) These are independent of the dataset.

3) The values of hyperparameters can be estimated by hyperparameter tuning such as entropy ,gini impurity.

Q13:What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?

Ans:Model based learning algorithm search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function. The model makes prediction by using the value of the new instance and the parameters in its function.

Q14: Can you name four of the most important Machine Learning challenges?

Ans:The challenges in Machine Learning includes overfitting the data, underfitting the data (using a simple model), missing data and non labelled data.

Q15: What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?

Ans: If the model performs too well on the training data but fails to generalize it can be due to overfiiting of the data.The three options can be look for outliers (if any),Go for a simpler model,and lastly try to get more data .

Q16:What exactly is a test set, and why would you need one?

Ans:Test set is a set that we test our model (fit using training data) to see how it performs. Test set is necessary so that we can determine how good (or bad) the model performs.

Q17:What is a validation set’s purpose?

Ans:The purpose of validation set is to basically compare different different models.

Q18:What precisely is the train-dev kit, when will you need it, how do you put it to use?

Ans: The goal of **dev-set** is to rank the models in term of their accuracy and helps us decide which model we need to proceed further with. Using Dev set we rank all our models in terms of their accuracy and pick the best performing model, in other words we can say that it basically filters out bad models and selects the best one.

Q19:What could go wrong if you use the test set to tune hyperparameters?

Ans:If we tune hyperparameters using the test set, then it may not perform well on the out-of-sample data and we won’t be able to get the generalized model.