ML Assignment Ineuron

Created	@May 18, 2021 4:30 PM
Tags	

1. What does one mean by the term "machine learning"?

When a program learns from experience on a particular task measured by a performance and gets better with the experience.

2.Can you think of 4 distinct types of issues where it shines?

Problems involving large amounts of data

Problems which are complex where the number of features are high machine learning helps

to make the solution simpler.

Problem statements involving constant updates of data and need to adapt continuously and

rapidly.

Automation of processes

3.What is a labeled training set, and how does it work?

A labeled training set contains the outputs of the data, this is used to train the model, the labels

to measure the performance of the model and tune the parameters.

4.What are the two most important tasks that are supervised?

Regression and Classification

5.Can you think of four examples of unsupervised tasks?

Clustering, Visualization & Dimensionality reduction, Anomaly and novelty detection, Associate

rule learning

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

Reinforcement Learning

7.Which algorithm will you use to divide your customers into different groups?

K-Means

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

Supervised

9.What is the concept of an online learning system?

Online learning is a system in which the system is trained incrementally by feeding sequences of data individually or in mini-batches. This kind of system doesn't need huge storage as the previously trained data can be wiped off and it is also fast and cheap.

10.What is out-of-core learning, and how does it differ from core learning?

When the dataset is so huge that it cannot fit in one machine's main memory then online learning system can be used to load a part of dataset and learn a step then repeat this process until we exhaust all the dataset this is called out-of-core learning whereas in core learning the machine can hold the entire dataset.

11.What kind of learning algorithm makes predictions using a similarity measure?

Instance based learning

12.What's the difference between a model parameter and a hyperparameter in a learning algorithm?

Model parameters are estimated on the data during model training and they are used by the model to make predictions. eg(weights in linear regression) whereas the model hyperparameters are set manually and are used in process to set the model parameters.

13. What do model based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?

Model based learning algorithms search for model parameters. To find the model parameters one can define either the utility function or the cost function, the most common one used is the cost function in which the functions measures how bad our predictions are. The predictions are made by

feeding in the feature values in our hypothesis function.

14.Can you name four of the most important Machine Learning challenges?

Insufficient data: Even more simpler algorithms perform identical to the complex algorithms when a sufficient amount of data is provided so lack of data may cause underperformance.

Non-representative data: If there is a lot of sampling data in our training data then we get a model which doesn't generalize well and hence will not perform well on the test data.

Over-fitting: There is the possibility that the model overfits the data which gives low bias but the problem is it increases the variance or in simpler words it doesn't generalize well and the exact converse happens when a model underfits the data i.e the model is not complex enough.

Poor Quality data: Presence of outliers, missing values, errors, noise causes in a model which doesn't perform well.

15.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?

The model is likely overfitting the if this happens.

We can fix this by decreasing the complexity of the model by selecting one with fewer parameters, decreasing the attributes in training dataset or by constraining the model.

Gathering more training data.

Reducing the noise in training data(data preprocessing).

16) What is a test set and why would you want to use it?

We can check whether our model generalizes well or not by testing it out on new cases we do that by dividing our data in 2 parts train and test. We train our model on train set and test its performance on the test set.

17.What is a validation set's purpose?

We train our models on the train set but now we want to tune over parameters to increase the performance we can't do it on the test set because it will familiarize the model with the test instances and we won't get the actual performance of the model so we hold out a part of a part of the data as validation set to tune the hyperparameters.

18) What can go wrong if you tune hyperparameters using the test set?

The model will get familiarized with the test data and it won't give the actual performance i.e it won't generalize well.

19) What is cross-validation and why would you prefer it to a validation set?

In cross-validation the training set is divided into different subsets and different combinations of of subsets are used to form the models and the remaining parts are used to validate the hyperparameters.

Cross-validation is preferable to a validation set because it saves from wasting to much data in validation sets.