**SIVAKUMAR PALANIRAJAN**

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

Machine learning is basically when we are trying to teach machines to make intelligent decisions just like human beings.

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

For credit card defaulters prediction, for Loan approval prediction, for grouping customers and their interests, for movie rating prediction

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

Labelled training set is the data where the class of each data is mentioned through which our model can be trained.

It works by having specific class labels for each subset of values of the input features. The target feature’s classes are mentioned.

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

Regression tasks and Classification tasks

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

Generative modelling, Clustering, Dimensionality reduction, Anomaly detection

**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?**

Clustering

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

Supervised learning problem as there are already two specific class ( Spam or not spam) which is basically a classification task

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

It is a different type of learning system where the system continues to learn again and again as new data is available.

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

Out-of-core learning is a method used in machine learning to handle datasets that are too large to fit into RAM. It processes the data in chunks that are small enough to fit into memory, allowing the algorithm to work with data that exceeds the system's RAM capacity. This contrasts with core learning, which assumes the entire dataset can be loaded into memory for processing.

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

K-Nearest Neighbors tends to use similarity measure

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

Model parameter is a parameter which tends to change throughout the training process of a model whereas hyperparameter is fixed right before the training starts and doesn’t change.

**13.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?**

The criteria include: good fit, good generalization and an appropriate complexity.

Most popular method: Gradient descent

Method to make predictions: Linear regression, logistic regression and so on

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

Less data quantity

Less data quality

Overfitting

Underfitting

**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?**

It is a basic case of overfitting where the model tends to perform well on training data but performs bad in testing as there is poor generalization.

We could fix it by: 1)Reducing testing error by increasing quality of data

2)Increase generalization by increasing quantity of data

3)Using regularization techniques like L1 and L2 regularization

**16.What exactly is a test set, and why would you need one?**

Test set is the other half of the data which is being used to test the model. Initially the whole data is divided into training and test data. Training data is used to train the model but we need some data which is different from the training data to test and analyse performance of the model so for testing, we use a test set.

**17.What is a validation set's purpose?**

Validation set is used to help fine tune the hyperparameters of the model to design the best model possible.

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

The train-dev set is an additional dataset used to ensure that the model's performance is not overly optimistic due to potential overfitting to the validation set. It serves as a middle ground between the training set, validation set, and test set.

We need it when we suspect overfitting.

After we tuning the hyperparameters using validation set, we use train-dev set to compare the performances between the model’s performance on validation set and train-dev set.

**19.What could go wrong if you use the test set to tune hyperparameters?**

If we use test set to tune hyperparameters then the model would be able to fine tune its hyperparameters to perfectly fit the test set which will lead to poor generalization. As we have tuned the hyperparameters using test set, the model will perform well in testing set but it isn’t reliable.