

Name:

Sejal Kaur Viridi

Subject Code:

18AI88

University Seat Number:

1DT20AI036

Date:

11-05-2024

Question 1

A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with Experience E .

3 examples are:-

→ Detecting tumor in scans:

This is semantic segmentation, where each pixel in the image is classified using CNN's.

→ Summarizing long documents:

Its a branch of NLP that summarizes long text.

→ Object Detection:

This is done using Image Segmentation using Deep Learning.

Question 2

The main challenges in Machine Learning is as follows:

→ Nonrepresentative Training Data:

It is necessary that the training data contains such samples as are needed for it to generalise. Otherwise it does not generalise well. If the sample is too small, we have sampling noise and large samples can be flawed. This is also called sampling bias.

→ Insufficient Quality of Training Data:

It usually takes lots of data for Machine Learning algorithms to work properly. Even for simple things a lot of examples are needed. Hence it is necessary to provide the ML algo with sufficient data.

→ Poor-Quality Data:

Data that contains errors, outliers and noise are considered to be poor Quality data. Such data makes it difficult for ML algorithms to identify underlying patterns. Therefore it is necessary to clean up data before using for Machine Learning.

→ Irrelevant Features:

A ML algorithm performs well only if it gets good features and not irrelevant ones. It is therefore necessary to come up with good set of training features. This is called feature engineering. Feature engineering involves:

Feature selection, feature extraction and creating new features.

→ Overfitting of Data:

This is when, the model performs well on training data but does not generalize the testing data well. This happens if the model is too complex. It can be reduced by simplifying the model and increasing training data.

→ Underfitting of Data:

This is when the model is unable to learn the underlying pattern. This happens when the model is too simple to predict. It can be reduced by selecting a more powerful model and feed better features.

Question 3

→ Voting Classifier:-

Suppose there are a few classifiers and each one achieves different accuracy. Suppose there is Logistic Regression, SVM, Decision Tree and so on. A better way to get the results with more accuracy is to aggregate the results of all classifiers and predict the class that get the most results. This is the hard-voting classifier. This method gives a better accuracy.

→ Bagging:

Either different classifiers can be used or predictors can be trained using same algorithm but train them on different random subsets of training set. When this sampling is performed with replacement, it is called Bagging. Only Bagging allows training instances to

be sampled several times for same predictor. After this aggregation can be applied to get the results.

→ Out of Bag Evaluation:

Some instances are sampled several times while some instances may not be sampled at all. A Bagging classifier by default samples only 'm' training instances with replacement. Since the rest are never seen to the predictor, they can be used for validation. This is called out-of-box evaluation.

Question4

The 3 major types of ML algorithms are:

→ Supervised:

Here the training data is labelled and the input is associated with an output. The model learns the relationship between input and output data enabling it to make predictions on new data.

→ Unsupervised:

Here the training algorithm learns with unlabelled data. The model learns patterns without guidance.

→ Reinforcement:

Here an agent learns to interact with the environment by taking actions and learning on feedbacks by rewards or penalties. Goal is to maximize rewards.

Question5

An example of supervised learning is classification of email as spam or not spam.

An example of unsupervised learning is using customer purchase history data to group them into segments.