**Assignment\_7**

1.What is the definition of a target function? In the sense of a real-life example, express the target function. How is a target function's fitness assessed?

**Ans: The target function is essentially the formula that an algorithm feeds data to in order to calculate predictions. As in algebra, it is common when training AI to find the variable from the solution, working in reverse.**

**Ex – If our network were to predict [cat,dog,cat,bird,dog], this would be called the “predicted array” which contains the predicted outputs.**

2.What are predictive models, and how do they work? What are descriptive types, and how do you use them? Examples of both types of models should be provided. Distinguish between these two forms of models.

**Ans: Predictive modelling is a commonly used statistical technique to predict future behaviour. Predictive modelling solutions are a form of data-mining technology tha works by analyzing historical and current data and generating a model to help predict future outcomes.**

**Descriptive studies can be several types, namely, case reports , case series, cross-sectional studies, and ecological studies. In the first three of these, data are collected on individuals, whereas the last one uses aggregated data for groups.**

**The difference between these two models is that descriptive analysis tells you what happened in the past. Predictive analytics predicts what is nost like;y to happen in the future**.

3.Describe the method of assessing a classification model's efficiency in detail. Describe the various measurement parameters.

**Ans: Logarithmic loss ( or log loss) measures the performance of a classification model where the prediction is a probability value between 0 and 1. Log loss increases as the predicted probability diverge from the actual label.**

**Parameters: Precision, Recall and Specificity, which are three major performance metrics describing a predictive classification model.**

4.

i. In the sense of machine learning models, what is underfitting? What is the most common reason for underfitting?

**Ans: When a model has not learned the patterns in the training data well and is unable to generalize well on the new data, it is known as underfitting. An underfit model has poor performance on the training data and will result in unreliable predictions. Underfitting occurs due to high bias and low variance.**

ii. What does it mean to overfit? When is it going to happen?

**Ans: Overfitting is a concept in data science, which occurs when a statistical model fits exactly against its training data. When this happens, the algorithm unfortunately cannot perform accurately against unseen data, defeating its purpose.**

iii. In the sense of model fitting, explain the bias-variance trade-off.

**Ans: In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.**

5. Is it possible to boost the efficiency of a learning model? If so, please clarify how.

**Ans: More often than not, machine learning models suffer from overfitting and their performance can be improved by using more training data.**

6. How would you rate an unsupervised learning model's success? What are the most common success indicators for an unsupervised learning model?

**Ans: Separation between two clusters can be computed by summating the distance between each pair of records falling within the two clusters and both the records are from different clusters.  
...  
Few examples of such measures are:**

**Silhouette coefficient.**

**Calisnki-Harabasz coefficient.**

**Dunn index.**

**Xie-Beni score.**

**Hartigan index.**

7. Is it possible to use a classification model for numerical data or a regression model for categorical data with a classification model? Explain your answer.

Ans:

8. Describe the predictive modeling method for numerical values. What distinguishes it from categorical predictive modeling?

**Ans: Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes.**

9. The following data were collected when using a classification model to predict the malignancy of a group of patients' tumors:

i. Accurate estimates – 15 cancerous, 75 benign

ii. Wrong predictions – 3 cancerous, 7 benign

Determine the model's error rate, Kappa value, sensitivity, precision, and F-measure.

10. Make quick notes on:

1. The process of holding out :

**Ans: The hold-out method for training a machine learning model is the process of splitting the data in different splits and using one split for training the model and other splits for validating and testing the models.**

2. Cross-validation by tenfold :

**Ans: With this method we have one data set which we divide randomly into 10 parts. We use 9 of those parts for training and reserve one tenth for testing. We repeat this procedure 10 times each time reserving a different tenth for testing**.

3. Adjusting the parameters:

**Ans: Parameters in machine learning and deep learning are the values your learning algorithm can change independently as it learns and these values are affected by the choice of hyperparameters you provide.**

11. Define the following terms:

1. Purity vs. Silhouette width :

**Silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually.**

**Purity is a measure of the extent to which clusters contain a single class. Its calculation can be thought of as follows: For each cluster, count the number of data points from the most common class in said cluster**

2. Boosting vs. Bagging :

**Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data.**

**Boosting is an iterative technique which adjusts the weight of an observation based on the last classification.**

3. The eager learner vs. the lazy learner :

**A lazy learner delays abstracting from the data until it is asked to make a prediction while an eager learner abstracts away from the data during training and uses this abstraction to make predictions rather than directly compare queries with instances in the datase**