**Machine Learning 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-) A target function in machine learning is a function that the model is trying to approximate or learn. For example, in a spam detection model, the target function is to distinguish between spam and non-spam emails. The fitness of a target function is evaluated by comparing the model's predictions with the actual labels in the training data.

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 models aim to predict an outcome based on input data, while descriptive models aim to describe the relationship between variables in the data. Examples of predictive models include linear regression and decision trees, while descriptive models include clustering and principal component analysis.

3. Describe the method of assessing a classification model’s efficiency in detail. Describe the various

measurement parameters.

Ans-) Classification model efficiency can be assessed using various metrics such as accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC).

4.Explain-

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

reason for underfitting?

Ans-) Underfitting occurs when the model is too simple and fails to capture the underlying patterns in the data. The most common reason for underfitting is the lack of sufficient training data or using an inappropriate model.

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

Ans-) Overfitting occurs when the model is too complex and fits the noise in the data instead of the underlying patterns. It can happen when the model is trained for too long, or when there is noise or outliers in the data.

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

Ans-) The bias-variance trade-off refers to the trade-off between a model's ability to fit the training data (low bias) and its ability to generalize to new data (low variance)

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

Ans-) Yes, the efficiency of a learning model can be boosted by using techniques such as feature engineering, ensemble methods, regularization, and hyperparameter tuning.

6. How would you rate an unsupervised learning model&#39;s success? What are the most common

success indicators for an unsupervised learning model?

Ans-) Unsupervised learning models can be evaluated using metrics such as within-cluster sum of squares, silhouette score, and purity.

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-) No, a classification model is designed to predict categorical outcomes, while a regression model is designed to predict numerical outcomes. Using the wrong type of model will lead to inaccurate predictions.

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

categorical predictive modeling?

Ans-) Predictive modeling for numerical values involves fitting a function to the input data that maps the input variables to a numerical output. It differs from categorical predictive modeling, which involves predicting the probability of each category.

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.

Ans-)

* Error rate = (3+7)/(15+75+3+7) = 10/100 = 0.1
* Kappa value = (15+7)/(15+3+7+75) - (18/100)\*(22/100) = 0.38
* Sensitivity = 15/(15+3) = 0.83 Precision = 15/(15+7) = 0.68
* F-measure = 2 \* (0.83 \* 0.68) / (0.83 + 0.68) = 0.75

10. Make quick notes on:

1. **The process of holding out**

Ans-) The process of holding out involves reserving a portion of the data for testing the model's performance after training.

2. **Cross-validation by tenfold**

Ans-) Cross-validation by tenfold involves dividing the data into ten equal parts, using nine parts for training and one part for testing, and repeating the process ten times.

3. **Adjusting the parameters**

Ans-) Adjusting the parameters involves tuning the model's hyperparameters to optimize its performance on the validation set.

11. Define the following terms:

1. **Purity vs. Silhouette width**

Ans-) Purity measures how well the clusters in a clustering model are separated, while silhouette width measures the average distance between the points in a cluster and the points in the neighboring clusters.

2. **Boosting vs. Bagging**

Ans-) Boosting and bagging are ensemble methods for combining multiple models to improve performance.

3. **The eager learner vs. the lazy learner**

Ans-) Eager learners build a model before receiving the test data, while lazy learners wait until the last minute to build the model, based on the test data.