**ASSIGNMENT**

**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&#39;s fitness assessed?**

**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.**

**3. Describe the method of assessing a classification model&#39;s efficiency in detail. Describe the various**

**measurement parameters.**

**4.**

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

**reason for underfitting?**

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

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

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

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

**success indicators for an unsupervised learning model?**

**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.**

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

**categorical predictive modeling?**

**9. The following data were collected when using a classification model to predict the malignancy of a**

**group of patients&#39; tumors:**

**i. Accurate estimates – 15 cancerous, 75 benign**

**ii. Wrong predictions – 3 cancerous, 7 benign**

**Determine the model&#39;s error rate, Kappa value, sensitivity, precision, and F-measure.**

**10. Make quick notes on:**

**1. The process of holding out**

**2. Cross-validation by tenfold**

**3. Adjusting the parameters**

**11. Define the following terms:**

**1. Purity vs. Silhouette width**

**2. Boosting vs. Bagging**

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

***SOLUTIONS***

1. ***A target function is a function that a machine learning model seeks to optimize or learn from. It maps input variables to output variables and can be used to make predictions. For example, a target function for a spam email filter might be a function that takes the email content as input and outputs a binary value indicating whether the email is spam or not. A target function's fitness is assessed by measuring how well it performs on a given task, using metrics such as accuracy, precision, recall, and F1 score.***
2. ***Predictive models are machine learning models that are used to make predictions about future or unseen data based on patterns in past data. They work by learning from historical data and using that knowledge to make predictions about new data. Descriptive models, on the other hand, are used to describe or summarize data and identify patterns or relationships within it. They are typically used for exploratory data analysis or data visualization. Examples of predictive models include linear regression, decision trees, and neural networks, while examples of descriptive models include clustering and principal component analysis.***
3. ***The efficiency of a classification model can be assessed using various measurement parameters such as accuracy, precision, recall, F1 score, and confusion matrix. Accuracy is the proportion of correct predictions made by the model, precision is the proportion of true positive predictions among all positive predictions, recall is the proportion of true positive predictions among all actual positives, F1 score is the harmonic mean of precision and recall, and a confusion matrix provides a breakdown of the model's predictions by actual class.***
4. ***i. Underfitting occurs when a machine learning model is too simple and fails to capture the complexity of the data. It usually happens when there are too few features or when the model is not flexible enough. ii. Overfitting occurs when a model is too complex and fits the training data too closely, resulting in poor performance on new or unseen data. It can happen when there are too many features or when the model is too flexible. iii. The bias-variance trade-off is a fundamental concept in model fitting that describes the trade-off between a model's ability to fit the training data (bias) and its ability to generalize to new data (variance). A model with high bias may underfit the data, while a model with high variance may overfit the data.***
5. ***Yes, it is possible to boost the efficiency of a learning model by optimizing its hyperparameters, increasing the size of the training data, reducing the dimensionality of the data, improving the quality of the input data, and selecting the appropriate algorithm for the problem at hand.***
6. ***The success of an unsupervised learning model is rated based on the evaluation of its clustering quality, dimensionality reduction, feature extraction, and anomaly detection performance. Common success indicators for unsupervised learning models include silhouette score, clustering error, variance explained, and reconstruction error.***
7. ***No, it is not appropriate to use a classification model for numerical data or a regression model for categorical data as these models are designed to work with specific types of data. Classification models are suitable for predicting categorical outcomes, while regression models are appropriate for predicting numerical outcomes.***
8. ***Predictive modeling for numerical values involves developing a mathematical model that can predict the numerical outcome of a particular phenomenon based on input variables. This is different from categorical predictive modeling, which involves predicting categorical outcomes based on input variables.***
9. ***Error rate = (3+7)/(15+75+3+7) = 10/100 = 0.1 or 10% Kappa value = 2\*((1575)-(37))/((15+7)(75+3)+(15+3)(75+7)) = 0.76 Sensitivity = 15/(15+3) = 0.83 or 83% Precision = 15/(15+7) = 0.68 or 68% F-measure = 2\*((0.68\*0.83)/(0.68+0.83)) = 0.75 or 75%***
   1. ***The process of holding out involves splitting the dataset into two subsets: one for training the model and the other for testing its performance.***
   2. ***Cross-validation by tenfold involves splitting the dataset into ten subsets and using each subset as a test set once while the rest are used for training.***
   3. ***Adjusting the parameters involves optimizing the hyperparameters of a model to improve its performance.***
   4. ***Purity measures how well a cluster contains only data points of the same class, while silhouette width measures how well the data points within a cluster are separated from data points in other clusters.***
   5. ***Boosting and bagging are two types of ensemble learning methods used to improve the performance of learning models by combining multiple models to make predictions.***
   6. ***Eager learners build a model using all the available data before making predictions, while lazy learners defer building the model until a prediction is required.***