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?

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'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'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' 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
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

Answer:

1. The target function, also known as the objective function or loss function, is a measure used to evaluate the accuracy of a machine learning model's predictions against the actual outcomes. A real-life example of a target function might be a model that predicts the price of a house based on various features such as the number of bedrooms, bathrooms, square footage, and location. The fitness of a target function is assessed by comparing the model's predicted output to the actual output using various performance metrics, such as accuracy, precision, recall, F1-score, or mean squared error, and then optimizing the function to improve the model's performance.
2. Predictive models are used to predict an output variable based on one or more input variables. They work by training a model on a set of input/output examples and then using that model to make predictions on new, unseen data. Descriptive models, on the other hand, are used to summarize and describe a dataset or phenomenon without necessarily making predictions. An example of a predictive model is a spam filter that predicts whether an incoming email is spam or not based on the email's content and metadata. An example of a descriptive model is a clustering algorithm that groups similar data points together based on their features. The key difference between these two types of models is that predictive models aim to make accurate predictions on new, unseen data, while descriptive models focus on understanding and summarizing the characteristics of the data.
3. The efficiency of a classification model is typically evaluated using various performance metrics, such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve. Accuracy is the ratio of correct predictions to the total number of predictions, while precision is the ratio of true positive predictions to the total number of positive predictions. Recall is the ratio of true positive predictions to the total number of actual positive cases, and F1-score is the harmonic mean of precision and recall. The ROC curve plots the true positive rate against the false positive rate for different threshold values, and the area under the curve (AUC) provides an overall measure of the model's performance.
4. i. Underfitting occurs when a model is too simple to capture the complexity of the data and results in high bias and low variance. The most common reason for underfitting is using a linear model to fit a non-linear relationship between the input and output variables. ii. Overfitting occurs when a model is too complex and fits the noise in the training data, resulting in low bias and high variance. Overfitting is likely to occur when a model has too many parameters relative to the number of training examples, or when the training data is noisy or contains outliers. iii. 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, unseen data (low variance). A model with high bias will underfit the data, while a model with high variance will overfit the data. The goal is to find a model with an appropriate balance between bias and variance, which can be achieved by adjusting the model's complexity or regularization.
5. Yes, there are several ways to improve the efficiency of a learning model, such as increasing the size or quality of the training data, selecting a more appropriate algorithm or hyperparameters, applying regularization techniques to prevent overfitting, or using ensemble methods to combine multiple models. Another approach is to perform feature engineering, which involves selecting or transforming the input features to improve the model's performance.