Q1.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?

Definition of a Target Function and Fitness Assessment:

* A target function is the ideal relationship between input and output in a machine learning problem.
* Example: In predicting house prices, the target function would be a mapping between features (e.g., size, location) and the actual price.
* Fitness is assessed by measuring how closely the model's predictions match the actual outcomes, often using metrics like Mean Squared Error (MSE) or accuracy.

Q2. 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.

* Predictive models make predictions about future outcomes based on input features (e.g., predicting stock prices).
* Descriptive models summarize patterns and relationships in data, providing insights (e.g., customer segmentation).
* Example of Predictive: Linear Regression for house price prediction.
* Example of Descriptive: K-Means Clustering for customer segmentation.

Q3. Describe the method of assessing a classification model's efficiency in detail. Describe the various measurement parameters.

Evaluate using metrics: Accuracy, Precision, Recall (Sensitivity), F1-Score, ROC Curve, AUC.

Q 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.

* Underfitting: Model is too simple to capture underlying patterns. Common reason: Model complexity is low.
* Overfitting: Model is too complex and fits noise in data. Happens with high model complexity.
* Bias-Variance Trade-off: Balance between underfitting (high bias) and overfitting (high variance).

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

Yes, by using techniques like ensemble methods (e.g., Gradient Boosting), fine-tuning hyperparameters, increasing data volume, and improving feature engineering.

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

Success indicators: Clustering accuracy, silhouette score, separation of clusters, interpretability of resulting groups.

Q 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.

Classification models are designed for categorical outcomes, while regression models are for numerical outcomes. Using a wrong model type may lead to inaccurate predictions.

Q8. Describe the predictive modeling method for numerical values. What distinguishes it from categorical predictive modeling?

* Predictive modeling for numerical values uses regression algorithms.
* For categorical values, classification algorithms are used.

Q 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.

* Error Rate: (Wrong Predictions) / (Total Predictions)
* Kappa Value: Calculation as described before.
* Sensitivity: (True Positives) / (True Positives + False Negatives)
* Precision: (True Positives) / (True Positives + False Positives)
* F-Measure: Combines Precision and Recall.

Q 10. Make quick notes on:

1. The process of holding out

2. Cross-validation by tenfold

3. Adjusting the parameters

* Process of Holding Out: Splitting data into training and testing sets.
* Cross-Validation by Tenfold: Dividing data into ten subsets for validation.
* Adjusting Parameters: Tuning model hyperparameters for optimal performance

Q 11. Define the following terms:

1. Purity vs. Silhouette width

2. Boosting vs. Bagging

3. The eager learner vs. the lazy learner

* Purity vs. Silhouette Width: Purity measures cluster homogeneity, Silhouette Width quantifies cluster separation.
* Boosting vs. Bagging: Boosting improves model sequentially, Bagging trains models in parallel.
* Eager Learner vs. Lazy Learner: Eager learners build a model during training, Lazy learners delay model creation until prediction is required.