1. Target Function:

Definition:

- Target Function: In machine learning, the target function represents the relationship between input features and the output variable that the model aims to learn.

Real-life Example:

- Example: Predicting house prices based on features like size, location, and number of bedrooms.

Fitness Assessment:

- Assessment: The fitness of a target function is evaluated by how well the model's predictions match the actual outcomes on new, unseen data.

2. Predictive vs. Descriptive Models:

Predictive Models:

- Definition: Predictive models make predictions about future outcomes based on historical data.

- Example: Linear regression predicting house prices.

Descriptive Models:

- Definition: Descriptive models summarize and describe patterns in data without making predictions.

- Example: Clustering algorithm grouping customers based on behavior.

Distinguishing Features:

- Focus:

- Predictive: Emphasizes making accurate predictions.

- Descriptive: Emphasizes summarizing patterns in data.

3. Classification Model Efficiency Assessment:

Assessment Method:

- Method: Evaluate using metrics like accuracy, precision, recall, F1 score, and area under the ROC curve.

- Parameters:

1. Accuracy: Overall correct predictions.

2. Precision: Accuracy of positive predictions.

3. Recall (Sensitivity): True positive rate.

4. F1 Score: Harmonic mean of precision and recall.

5. ROC Curve: Trade-off between true positive rate and false positive rate.

4. Underfitting, Overfitting, Bias-Variance Trade-off:

i. Underfitting:

- Definition: Model is too simple, fails to capture underlying patterns.

- Common Reason: Lack of complexity in the model.

ii. Overfitting:

- Definition: Model is too complex, fits noise in the data.

- When: Small datasets or overly complex models.

iii. Bias-Variance Trade-off:

- Trade-off: Balancing bias (error from too much simplicity) and variance (error from too much complexity) for optimal model performance.

5. Boosting Model Efficiency:

- Methods:

1. Feature Engineering: Improve input features.

2. Ensemble Methods: Combine multiple models for better predictions.

3. Hyperparameter Tuning: Optimize model parameters.

6. Unsupervised Learning Model Success:

- Success Indicators:

1. Cluster Cohesion: How well data points within a cluster stick together.

2. Cluster Separation: How distinct clusters are from each other.

7. Classification vs. Regression Models:

- Applicability:

- Classification Model: Used for predicting categorical outcomes.

- Regression Model: Used for predicting numerical values.

8. Predictive Modeling for Numerical vs. Categorical Values:

- Numerical Predictive Modeling:

- Example: Predicting house prices using regression.

- Distinguishing Factor: Outcome variable is continuous.

- Categorical Predictive Modeling:

- Example: Predicting customer churn using classification.

- Distinguishing Factor: Outcome variable is discrete categories.

9. Classification Model Metrics:

- Error Rate: Wrong Predictions\ Total Predictions = {3 + 7}\{15 + 75}

- Kappa Value, Sensitivity, Precision, F-Measure: Requires actual values for calculation.

10. Quick Notes:

1. Holding Out Process:

- Dividing the dataset into training and test sets for model evaluation.

2. Cross-Validation by Tenfold:

- Splitting data into 10 folds, training on 9 and validating on 1, repeating 10 times.

3. Adjusting Parameters:

- Fine-tuning model parameters for better performance.

11. Term Definitions:

1. Purity vs. Silhouette Width:

- Purity: Measure of homogeneity within clusters in unsupervised learning.

- Silhouette Width: Measures how well-separated clusters are.

2. Boosting vs. Bagging:

- Boosting: Sequentially builds models, focusing on errors of previous ones.

- Bagging: Parallelly builds models, combining them for more robust predictions.

3. Eager Learner vs. Lazy Learner:

- Eager Learner: Builds a generalized model during training (e.g., Decision Trees).

- Lazy Learner: Learns and generalizes during prediction (e.g., k-Nearest Neighbors).