

Assignment - 1 :- Basic of Machine Learning

Q-1 What is Hypothesis in Machine Learning?

→ A hypothesis is a function that maps input features to output predictions. It represents the model's assumptions about the data and is adjusted during training to minimize the errors.

Q-2 Explain the impact of taking the learning rate?

→ Smaller Learning Rate :- Leads to slow convergence but ensures stability and better accuracy. Suitable when data is complex or noisy.

→ Larger Learning Rate :- speeds up convergence but may overshoot the optimal point, risking divergence. Suitable for simpler datasets.

Q-3 What is overfitting, and how can you avoid it?

→ Overfitting occurs when a model performs

Well on training data but poorly on unseen data.

→ Prevention Techniques:

- Cross-validation
- Pruning for decision trees
- Regularization (L1, L2)
- Early stopping
- Increasing training data
- Reducing model complexity

Q-4 Why is Linear Regression Not Preferred For Classification?

→ Linear Regression predicts continuous values and lacks clear decision boundaries, making it ineffective for discrete class labels.

Logistic regression or decision trees are better suited for classification.

Q-5 Why Do We Perform Normalization?

→ Normalization scales features to a common range (e.g., 0-1), improving model convergence speed and accuracy. It's crucial for distance-based algorithms like K-NN, K-means and gradient descent.

Q-6 What is the Bias-Variance Tradeoff?

- High Bias :- Underfitting, model is too simple
- High Variance :- overfitting, model is too complex
- The tradeoff aims to find a balance for optimal performance

Q-7 What is a Training Set and Test Set?

- Training Set :- Used to train the model.
- Test set :- Used to evaluate model performance
- Validation set :- Helps tune hyperparameters to improve performance

Q-8 How to Handle Missing or Corrupted Data?

- Imputation :- Filling missing values using mean, median, or mode
- Deletion :- Remove rows or columns with excessive missing data
- Prediction Models :- Use models to predict missing values
- Flagging :- Add indicators for missing values

Q-9 Evaluation Metrics For Classification

- Accuracy :- overall correct predictions.
- Precision :- Correct positive predictions out of total predicted positives.
- Recall :- Correct positive predictions out of actual positives.
- F1 Score :- Harmonic mean of precision and recall.
- ROC - AUC :- Evaluates true positive vs false positive rates.

Q-10 Evaluation Metrics For Regression

- Mean Absolute Error (MAE) :- Average absolute difference between predicted and actual values.
- Mean Squared Error (MSE) :- Average squared difference; penalizes large errors.
- Root Mean Squared Error (RMSE) :- Square root of MSE; interpretable in original units.
- R^2 Coefficient of Determination :- Measures variance explained by the model.