Machine Learning Interview Questions & Answers (PDF Notes)

Q: What is Machine Learning?

A: Machine Learning is a subset of AI where systems learn patterns from data and improve performance without being explicitly programmed.

Q: Difference between supervised and unsupervised learning?

A: Supervised learning uses labeled data, while unsupervised learning works on unlabeled data to find hidden patterns.

Q: What is overfitting?

A: When a model performs very well on training data but poorly on unseen test data.

Q: What is underfitting?

A: When a model is too simple to capture underlying patterns in data.

Q: What is bias-variance tradeoff?

A: It is the balance between error from wrong assumptions (bias) and sensitivity to training data (variance).

Q: What is cross-validation?

A: A technique to evaluate model performance by splitting data into multiple train-test folds.

Q: What is a confusion matrix?

A: A table showing True Positives, False Positives, True Negatives, and False Negatives.

Q: Explain precision and recall.

A: Precision measures correctness of positive predictions. Recall measures how many actual positives were captured.

Q: What is regularization?

A: A technique to prevent overfitting by penalizing large model weights.

Q: Difference between L1 and L2 regularization?

A: L1 can make weights zero (feature selection). L2 shrinks weights but does not make them zero.

Q: What is Gradient Descent?

A: An optimization algorithm used to minimize loss function by updating parameters iteratively.

Q: What is learning rate?

A: It controls how big the step size is in gradient descent.

Q: What is a hyperparameter?

A: A value set before training that controls model behavior, like max_depth or learning_rate.

Q: What is a parameter?

A: A value learned from data during training, like weights in linear regression.

Q: What is Random Forest?

A: An ensemble method using multiple decision trees with bagging and voting/averaging.

Q: What is boosting?

A: A technique where models are trained sequentially to correct previous errors.

Q: What is PCA?

A: A dimensionality reduction technique that projects data into principal components.

Q: What is data leakage?

A: When information from test/future data is accidentally used in training.

Q: What is model drift?

A: When real-world data distribution changes over time, reducing model performance.

Q: What is inference?

A: The process of using a trained model to make predictions on new data.