

1. What is the difference between supervised and unsupervised learning?

Answer :

Supervised Learning:

- Uses labeled data to train the model.
- The model learns the mapping between input and output.
- Examples: Regression, Classification.

## Unsupervised Learning:

- Uses unlabeled data to find patterns or intrinsic structures.
- No specific output is predicted.
- Examples: Clustering, Dimensionality Reduction.

2. What is overfitting and how can you prevent it?

Answer :

Overfitting:

- Occurs when a model learns both the training data and the noise within it, performing well on training data but poorly on new, unseen data.

### Prevention Techniques:

- Cross-validation.
- Pruning in decision trees.
- Regularization (L1 and L2).
- Reducing the complexity of the model.

- Using more training data.
- Early stopping in iterative models.

3. Explain the bias-variance tradeoff.

Answer :

- Bias: Error introduced by approximating a real-world problem, which may be complex, by a simplified model. High bias can cause

underfitting.

- Variance: Error introduced due to the model's sensitivity to small fluctuations in the training set. High variance can cause overfitting.

- Tradeoff: A balance between bias and variance is essential for building a model that generalizes well to unseen data.

## 4. What is the purpose of A/B testing?

### Answer :

- A/B Testing:
  - A method to compare two versions of a variable to determine which one performs better.
  - Used to test changes to a webpage, app, or marketing campaign against the current

version.

- Helps in making data-driven decisions.

5. What is the difference between Type I and Type II errors?

Answer :

Type I Error:

- Also known as a false positive.
- Occurs when the null hypothesis is rejected

when it is actually true.

Type II Error:

- Also known as a false negative.
- Occurs when the null hypothesis is not rejected when it is actually false.

6. Explain the concept of cross-validation.

Answer :

Cross-Validation:



- A technique for assessing how the results of a statistical analysis will generalize to an independent data set.
- Common methods: K-Fold Cross-Validation, Leave-One-Out Cross-Validation.
- Helps in mitigating overfitting and selecting the best model.

7. What are some common metrics for evaluating the performance of a classification model?

Answer :

Common Metrics:

- Accuracy:  $(TP + TN) / (TP + TN + FP + FN)$ .
- Precision:  $TP / (TP + FP)$ .
- Recall:  $TP / (TP + FN)$ .

- F1 Score:  $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ .
- ROC-AUC: Area under the receiver operating characteristic curve.

8. What is a confusion matrix?

Answer :

Confusion Matrix:

- A table used to describe the performance of a

classification model.

- Comprises True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

9. Explain the difference

between bagging and boosting.

Answer :

Bagging (Bootstrap Aggregating):

- Reduces variance by training multiple models on different subsets of data and averaging their predictions.
- Example: Random Forest.

Boosting:

- Reduces bias by combining weak learners sequentially, each correcting the errors of its predecessor.
- Example: AdaBoost, Gradient Boosting.

10. What is dimensionality reduction and why is it important?

## Answer :

Dimensionality Reduction:

- The process of reducing the number of random variables under consideration.
- Important for:
  - Reducing computation time.
  - Removing multicollinearity.
  - Reducing noise and improving model performance.

- Visualization in 2D or 3D.

Follow for more  
informative content



**Amar Sharma** (He/Him)

AI Engineer at Horizon Broadband Pvt. Ltd. •  
ex Data Scientist at Rubixe | Machine Learning  
| Deep Learning | AWS | NLP | NER | GenAI |  
GAN | Vector Database | LLM | LangChain | AI  
Products Research Team Member

Horizon Broadband Private Limited  
Bengaluru, Karnataka, India

**15,854 followers** · **500+ connections**