

Chapter – 1

The Machine Learning Landscape

Exercise

1. How would you define Machine Learning?

A field of AI where systems learn from data to make decisions or predictions without being explicitly programmed.

2. Can you name four types of problems where it shines?

Classification, regression, clustering, and recommendation.

3. What is a labeled training set?

A dataset with input-output pairs where the correct output (label) is known for each input.

4. What are the two most common supervised tasks?

Classification and regression.

5. Can you name four common unsupervised tasks?

Clustering, anomaly detection, dimensionality reduction, and association rule learning.

6. What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?

Reinforcement Learning.

7. What type of algorithm would you use to segment your customers into multiple groups?

Clustering algorithm (e.g., K-Means).

8. Would you frame the problem of spam detection as a supervised learning problem or an unsupervised learning problem?

Supervised learning.

9. What is an online learning system?

A system that learns continuously from incoming data streams.

10. What is out-of-core learning?

Learning from data that doesn't fit into memory by using small data chunks.

11. What type of learning algorithm relies on a similarity measure to make predictions?

Instance-based learning (e.g., k-Nearest Neighbors).

12. What is the difference between a model parameter and a learning algorithm's hyperparameter?

Model parameters are learned from data; hyperparameters are set before training.

13. What do model-based learning algorithms search for?

What is the most common strategy they use to succeed?

How do they make predictions?

They search for the best model parameters using optimization (e.g., gradient descent). They use the learned model to make predictions.

14. Can you name four of the main challenges in Machine Learning?

Overfitting, underfitting, data quality, and interpretability.

15. If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name three possible solutions?

Overfitting. Possible solutions are more data, regularization, cross-validation.

16. What is a test set and why would you want to use it?

A dataset used to evaluate model performance on unseen data to assess generalization.

17. What is the purpose of a validation set?

To tune hyperparameters and evaluate model performance during training.

18. What can go wrong if you tune hyperparameters using the test set?

The model overfits the test set, giving misleading performance estimates.

19. What is repeated cross-validation and why prefer it over a single validation set?

It's cross-validation done multiple times with different splits; gives more reliable, averaged performance results.