

# Machine Learning Cheatsheet & MCQs

---

## Cheatsheet

1. **Definition of Machine Learning**
  - Machine Learning is a field of AI that enables systems to learn from data and improve performance without being explicitly programmed.
2. **Four Types of Applications Where ML Shines**
  - Image and Speech Recognition
  - Fraud Detection
  - Recommendation Systems
  - Autonomous Vehicles
3. **Labeled Training Set**
  - A dataset where each example is paired with an output label.
4. **Two Most Common Supervised Tasks**
  - Regression
  - Classification
5. **Four Common Unsupervised Tasks**
  - Clustering
  - Anomaly Detection
  - Dimensionality Reduction
  - Association Rule Learning
6. **Algorithm for a Robot Walking in Unknown Terrains**
  - Reinforcement Learning Algorithm
7. **Algorithm for Customer Segmentation**
  - Clustering Algorithm (e.g., K-Means)
8. **Spam Detection Problem Type**
  - Supervised Learning Problem
9. **Online Learning System**
  - A model that updates continuously as new data arrives, without retraining on past data.
10. **Out-of-Core Learning**
  - Training a model using mini-batches when the dataset is too large to fit in memory.
11. **Algorithm Based on Similarity Measures**
  - Instance-Based Learning (e.g., K-Nearest Neighbors)
12. **Difference Between Model Parameter and Model Hyperparameter**
  - Model parameters are learned from data (e.g., weights in neural networks).
  - Hyperparameters are set before training (e.g., learning rate, number of layers).
13. **What Model-Based Algorithms Search For**
  - They search for the best parameters to minimize the loss function and use optimization techniques like gradient descent.
14. **Four Main Challenges in ML**
  - Overfitting
  - Underfitting
  - Data Quality Issues
  - Scalability
15. **Poor Generalization Causes and Solutions**

- **Cause:** Overfitting
  - **Solutions:** More data, Regularization, Feature Selection
  - 16. **Test Set Purpose**
    - A dataset used to evaluate model performance on unseen data.
  - 17. **Validation Set Purpose**
    - Used to fine-tune hyperparameters and avoid overfitting.
  - 18. **Train-Dev Set Usage**
    - Used when the training set is large but different from test data, ensuring reliable tuning.
  - 19. **Risk of Tuning Hyperparameters with the Test Set**
    - Leads to overfitting to the test set and poor generalization.
- 

## Multiple Choice Questions (MCQs)

1. What is Machine Learning?
  - (A) A method to manually program systems
  - (B) A field of AI that learns from data (*Correct*)
  - (C) A way to write better algorithms
  - (D) A subset of Robotics
2. Which of the following is NOT an application of ML?
  - (A) Spam Detection
  - (B) Facial Recognition
  - (C) Word Processing (*Correct*)
  - (D) Fraud Detection
3. What is an example of a supervised learning task?
  - (A) Clustering
  - (B) Association Rule Learning
  - (C) Regression (*Correct*)
  - (D) Anomaly Detection
4. Which algorithm is best suited for customer segmentation?
  - (A) K-Means (*Correct*)
  - (B) Linear Regression
  - (C) Neural Networks
  - (D) Reinforcement Learning
5. What problem occurs when a model performs well on training data but poorly on new instances?
  - (A) Underfitting
  - (B) Overfitting (*Correct*)
  - (C) Optimal Learning
  - (D) Data Scaling
6. What is the purpose of a validation set?
  - (A) Training the model
  - (B) Evaluating model accuracy
  - (C) Tuning hyperparameters (*Correct*)
  - (D) Storing labeled data
7. Why shouldn't you tune hyperparameters on the test set?
  - (A) It's computationally expensive
  - (B) It causes underfitting

- (C) It leads to overfitting (*Correct*)
  - (D) It increases variance
8. Which of the following is an example of online learning?
- (A) Training a model once and using it forever
  - (B) Updating a model continuously with new data (*Correct*)
  - (C) Running a model on offline data
  - (D) Using pre-trained models
9. What is an Out-of-Core Learning strategy useful for?
- (A) Training large datasets that don't fit in memory (*Correct*)
  - (B) Making real-time predictions
  - (C) Increasing model interpretability
  - (D) Reducing overfitting
10. What is the main challenge in machine learning?
- (A) Data availability
  - (B) Choosing the right algorithm
  - (C) Overfitting (*Correct*)
  - (D) Writing efficient code