Chapter 1 Exercise Questions

1. How would you define Machine Learning?
   Machine Learning focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

2. Can you name four types of problems where it shines?
   - Classification
   - Regression
   - Anomaly Detection
   - Clustering

3. What is a labeled training set?
   Labelled training sets have the corresponding target or output value and is used in supervised machine learning algorithms.

4. What are the two most common supervised tasks?
   - Classification
   - Regression

5. Can you name four common unsupervised tasks?
   - Clustering
   - Dimensionality reduction
   - Anomaly detection
   - 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

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?
   An online learning system is a machine learning system that can continuously learn from incoming data.

10. What is out-of-core learning?
    Out-of-core learning refers to machine learning algorithms that can handle datasets that are too large to fit into the computer's main memory.

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

12. What is the difference between a model parameter and a learning algorithm's hyperparameter?
    Model parameters learn from model during training and learning algorithms hyperparameter learn from instance data.

13. What do model-based learning algorithms search for? What is the most common strategy they use to succeed. How do they make predictions?
    Model-based learning algorithms search for optimal value. They make predictions as they have got trained.

14. Can you name four of the main challenges in Machine Learning?
    - Overfitting
    - Underfitting
    - Data scarcity
    - Data preprocessing

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?
    Possible solutions include simplifying the model, increasing the size of the training data.

16. What is a test set, and why would you want to use it?
    A portion of dataset used for evaluating the model performance.

17. What is the purpose of a validation set?
    Tume hyperparameters and assess model performance.

18. What is the train-dev set, when do you need it, and how do you use it?
    The train-dev set is a subset of the training data that is further split from the validation set.

12. 19. What can go wrong if you tune hyperparameters using the test set?
    Tuning hyperparameters using the test set can lead to overfitting to the test data,