**Chapter 1: The Machine Learning Landscape**

1. ***How would you define machine learning?***

* Machine learning is the art and science of programming computers so that they can learn from data.
* Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.

1. ***Can you name 4 types of problems where it shines?***

* Spam filter
* Speech recognition
* Creation of chatbot or personal assistant
* Forecasting a company’s revenue for next year, based on many performance metrics

1. ***What is a labelled training set?***

* Each data point of the data set is marked with a label.
* For example, in a labelled training set of words describing emotions, “cheerful” and “victorious” may be labelled as positive whereas “angry” and “frustrated” may be labelled as negative.

1. ***What are the two most common supervised tasks?***

* *Classification:* Each data point falls into a certain ‘class’.
* *Regression:* Each data point is associated with a numerical value.

1. ***Can you name 4 most common unsupervised tasks?***

* Clustering
* Anomaly and Novelty Detection
* Visualisation and Dimensionality Reduction
* Associated Rule Learning

1. ***What type of machine learning algorithm would you use to allow a robot to walk in various unknown terrains?***

* *Reinforcement Learning:* A learning system (agent) that can observe the environment, select and perform actions and get rewards (or penalties) in return.

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

* *Clustering Unsupervised Algorithm:* This will help detect groups of similar customers.

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

* *Classification Supervised Algorithm:* The spam detection model is trained with labelled set of emails (spam or not spam) and then it helps to flag new emails.

1. ***What is an online learning system?***

* The system is trained by incrementally feeding it data instances sequentially, either individually or in mini-batches.
* Each learning batch is fast and cheap, so the system can learn about new data on the fly, as it arrives.

1. ***What is out-of-core learning?***

* Online learning algorithms can be used to train systems on huge datasets that cannot fit into one machine’s main memory.
* The algorithm loads part of the data, runs a training step on that data, and repeats the process until it has run on all of the data.
* It is usually done offline.

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

* *Instance-based learning algorithm:* The system uses a similarity measure to compare the new cases to the learned examples.

1. ***What is the difference between a model parameter and a learning algorithm’s hyperparameter?***

* A hyperparameter is a parameter of the learning algorithm, not the model.
* A model parameter is a parameter of the model.

1. ***What do model-based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?***
2. ***Can you name 4 of the main challenges in machine learning?***
3. ***If your model performs great on the training data but generalises poorly to new instances, what is happening? Can you name 3 possible solutions?***
4. ***What is a test set and why would you want to use it?***
5. ***What is the purpose of a validation set?***
6. ***What is the train-dev set, when do you need it? And how do you use it?***
7. ***What can go wrong if you use hyperparameters using the test set?***