**Exercises**

1. How would you define Machine Learning?

|  |
| --- |
| Machine Learning is the capability that a given system has to learn from data in order to give a response for a given purpose. |

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

|  |
| --- |
| 1. Mail spam categorization 2. Predict the value of a given house 3. Voice recognition 4. Text categorization |

1. What is a labeled training set?

|  |
| --- |
| A labelled training set is a set were we aready know the target (labels) the system has to predict. |

1. What are the two most common supervised tasks?

|  |
| --- |
| Classification and Regression |

1. Can you name four common unsupervised tasks?

|  |
| --- |
| Clustering, anomaly detection and novelty detection, Visualization and dimensionality reduction, Association rule learning. |

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

|  |
| --- |
| Reinforcement learning |

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

|  |
| --- |
| Clustering |

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

|  |
| --- |
| Supervised |

1. What is an online learning system?

|  |
| --- |
| A system that uses a given algorithm to learn from data. It could be instance or model-based |

1. What is out-of-core learning?

|  |
| --- |
| Out-of-core learning is performed when a given environment has not the necessary resources to make out model learn, the data is split into chunks and the algorithm learns continuously. |

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

|  |
| --- |
| Instance-based learning algorithm |

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

|  |
| --- |
| A model parameter is any type of the data that we use to make predictions. An hyperparameter is a parameter used to fine tune the decision the algorithm takes. It makes the model to fit more/less on data depending on its value. |

1. 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 patterns. Formulas. … |

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

|  |
| --- |
| 1. Insufficient quantity of training data 2. Non-representative training data 3. Poor-quality data 4. Irrelevant features 5. Over/under fitting the training data |

1. 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.  Solutions:   * Simplify the model by selecting one with fewer parameters (e.g., a linear model rather than a high-degree polynomial model), by reducing the number of attributes in the training data, or by constraining the model. * Gather more training data. * Reduce the noise in the training data (e.g., fix data errors and remove outliers). |

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

|  |
| --- |
| The test set is used to test our model. We use it to fine-tune the hyperparameters and to check how well our model performs on new data. |

1. What is the purpose of a validation set?

|  |
| --- |
| To fine tune the hyperparameters before testing it with the test set in order to avoid establishing the values of the hyperparameters accordingly to the specific test set. |

1. What is the train-dev set, when do you need it, and how do you use it?

|  |
| --- |
| After the model is trained (on the training set, *not* on the train-dev set), you can evaluate it on the train-dev set |

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

|  |
| --- |
| The problem is that you measured the generalization error multiple times on the test set, and you adapted the model and hyperparameters to produce the best model *for that particular set*. This means that the model is unlikely to perform as well on new data. |