1. Machine learning is **the concept that a computer program can learn and adapt to new data without human intervention**. Machine learning is a field of artificial intelligence (AI) that keeps a computer's built-in algorithms current regardless of changes in the worldwide economy.
2. **Machine learning algorithms** have had good results on problems such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.
3. The training set is used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known. The final step is to compare the predicted responses against the actual (observed) responses to see how close they are. In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it.
4. Supervised machine learning tasks can be broadly classified into two subgroups: **regression and classification**. Regression is the problem of estimating or predicting a continuous quantity.
5. Common unsupervised tasks include **clustering, visualization, dimensionality reduction, and association rule learning**.
6. The best Machine Learning algorithm to allow a robot to walk in unknown terrain is **Reinforced Learning**, where the robot can learn from response of the terrain to optimize itself.
7. **K-Means clustering** is an unsupervised machine learning algorithm that divides the given data into the given number of clusters.
8. **Spam detection is a supervised machine learning problem**. This means you must provide your machine learning model with a set of examples of spam and ham messages and let it find the relevant patterns that separate the two different categories.
9. online machine learning is **a method of machine learning in which data becomes available in a sequential order and is used to update the best predictor for future data at each step**, as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set
10. Out-of-core learning refers to **a set of algorithms working with data that cannot fit into the memory of a single computer**, but that can easily fit into some data storage such as a local hard disk or web repository.
11. Learning algorithm that relies on a similarity measure to make predictions is **instance-based algorithm**.
12. **Model parameters are estimated based on the data during model training and model hyperparameters are set manually and are used in processes to help estimate model parameters**.
13. Four main challenges in Machine Learning include **overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data**.
14. Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.Three different options can be cross-validation,regularization,and train with more data.
15. A test set in machine learning is **a secondary (or tertiary) data set that is used to test a machine learning program after it has been trained on an initial training data set**.
16. validation set is a set of data used to train artificial intelligence (AI) with the goal of **finding and optimizing the best model to solve a given problem**.
17. The goal of dev-set is **to rank the models in term of their accuracy and helps us decide which model to proceed further with**. Using Dev set we rank all our models in terms of their accuracy and pick the best performing model.
18. lack of data,

poor data quality,

nonrepresentative data,

uninformative features,

excessively simple models that underfit the training data.

complex models that overfit the data.