**Homework 1**

**By Madhumita Kathuria**

Q1) How would you define Machine Learning?

Ans: Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

2) What are the differences between Supervised and Unsupervised Learning? Specify example 3 algorithms for each of these.

Ans: A supervised learning algorithm learns from labelled training data, helps us to predict outcomes for unforeseen data.

Algorithms:

* Linear regression
* KNN
* Random Forest
* SVM

Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data. Unsupervised learning algorithms allow you to perform more complex processing tasks compared to supervised learning.

Algorithms:

* Hierarchical clustering
* K-means clustering
* K-NN (k nearest neighbors)
* Principal Component Analysis
* Singular Value Decomposition
* Independent Component Analysis

3) What are the test and validation set, and why would you want to use them?

Ans: That the validation dataset is used to describe the evaluation of models when tuning hyperparameters and data preparation. We, as machine learning engineers, use this data to fine-tune the model hyperparameters.

The “test dataset” is used to describe the evaluation of a final tuned model when comparing it to other final models. It is only used once a model is completely trained(using the train and validation sets).

4) What are the main pre-processing steps? Explain them in detail. Why we need to prepare our data?

Ans:

* Acquire the dataset
* Import all the crucial libraries
* Import the dataset
* Identifying and handling the missing values
* Encoding the categorical data
* Splitting the dataset
* Feature scaling

5) How you can explore continuous and discrete variables?

Ans: **Discrete variables**are countable in a finite amount of time. For example, you can count the change in your pocket. You can count the money in your bank account.

Continuous variables, on the other hand, are defined as numbers or a numeric date that can take on any value. A numeric date means, for example, 01/01/2019 as opposed to names for days of the week or for months. Continuous variables are uncountable. For example, height is **not a countable** variable.

6) Analyse the plot given below. (What is the plot and variable type, check the distribution and make comment about how you can preproccess it.)

In [1]:

Ans:

Bar plot

Discrete variable

Standard distributions

Out[1]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x2b8ced01320>

