Introduction

ML Lifecycle :

You are familiar with definition of ML, that 'it enables machines to learn from data and predict outcomes on unseen data'. You are also familiar with terms like model, training / learning, data, etc. Now let's discuss the life cycle of ML, this is generally followed, in some cases this may not be followed. So here are the steps in life cycle of solving a problem through ML.

1. Define the problem : Define the problem statement clearly, you must be very clear about the problem, the problem statement, I/O, rough idea about technique (like regression, classification, etc.), rough idea about data (image, text, etc.) must be very clear in your mind. You must have idea on previous work done by others on the problem.

2. Get the data : Get the dataset for that problem. The best platform to get dataset is Kaggle, but in all cases you will not find data from there. You may find data from GitHub and other websites also. Explore various datasets and get the best dataset for your problem.

3. Clean the data : Once you have got the dataset, now your job is to 'clean' it means, make it suitable for training. The data can be in various forms, image, text, numerical, video, audio, etc. You need to perform data preprocessing before using the data. Sometimes you need to change the form of data i.e., audio to image.

4. Split the data : Before feeding to the model, you need to split the data into train and test. Once you have split the data, don't even look at test data. Use test data only when you come with the final strategy. Whatever you do, do it with train data. Sometimes you don't get sufficient data, you need to perform data augmentation step to get variety of data.

5. Select the model : Based on your problem select the model. The model selection is a tricky task which can be learnt by experience. But many times by observing data and the problem you can get insight about 3-4 best performing models. Use them in combined or select one from them.

6. Tune the model : Once you have selected the model, tune it. Find its best hyper parameters (we will see this terms) that meets perfectly to the output. This is also tricky step, you will learn by experience.

7. Train and predict : Train the model, on the data, and once the model is trained, predict the outcomes. In deep learning training the model also requires much attention like early stopping and other methods are used. That we will see in deep learning section.

8. Use the model : Once you have a trained model, use it on new data. Deploy it on various platforms. And get best outcomes.

Types of ML systems :

ML systems can be categorised based on three types such as

1. Supervision based - includes supervised, unsupervised, semi supervised and reinforcement
2. Data stream based - batch (offline) learning and online learning
3. Learning based - instance based and model based

Supervision based :

Supervision based ML systems differ on supervision they are provided in training. Here supervision means providing labelled data. There are four types belong to this category,

1. Supervised learning : As its name suggests, the model learns under supervision. It is provided the data with label. i.e., question and answer. The models learns the mapping between data and label and generalise to unseen data. Examples are - regression and classification models like linear regression, logistic regression, SVM, tree based algorithms

2. Unsupervised learning - Opposite to supervised learning, in this type data without label is provided to model. The model observes pattern from data and learns, it predicts outcome on unseen data. It is used in clustering, anomaly detection, association, etc.

3. Semi supervised learning - In this learning approach we provide labels to few data points. And the model learns by itself and labels the remaining datapoints. Consider google photos, we provide name of person (label) to single photo and the model itself recognises that all photos with that person. And labels those photos.

4. Reinforcement learning - In RL, the model learns by punishment or reward at each step in the environment. Terms - agent, environment, state, punishment / reward, policy. Examples - Robotics, chess, tic tac toe

Data stream based :

The model is trained based on data stream. Two types - online and offline (batch)

In offline, new model is trained on new dataset replacing the older model
In online, the same model is trained on new dataset, learning factor is considered here, LF tells about the influence of new instance on model's parameter

Learning based :

Instance based learning - remembers instances, KNN
Model based learning - finds parameters, linear regression