**6 Jars of the Machine learning**

Traditional program takes data and generates the output. But for Machine learning, machines are trained with the data and output to create a program or model.

In machine learning we (1) take some data, (2) train a model on that data, and (3) use the trained model to make predictions on new data. These three process steps can further categorized into 6 jars as given below:

1. Data

2. Tasks

3. Models

4. Loss function

5. Learning algorithm

6. Evaluation

1. **Data :**

Data can be input data and output data. To process the data using machine learning, this data should be in machine readable format. All data should be encoded as numbers typically high dimensional.

There are variety of sources available to avail the data like public websites/sources, paid sources and finally creating one’s own data (for testing purpose)

1. **Tasks :**

Once we get the data. What is the next step?. What do we do with the data?. Here task comes into the picture. For example, based your searches on any e commerce website, it shows some other products which you may like. The task is analyze the data and provide or suggest some other interested products.

We can divide tasks into:

1. Supervised learning tasks
2. Unsupervised learning tasks.

Supervised tasks: To perform supervised tasks, both input and out data i.e coordinates (x and y) are important. These are further categorized into:

Classification tasks: Learns relationship between coordinates x and y and output is 1 and zero Many of the real world tasks comes under classification tasks.

Regression tasks: These type of tasks neither interested in input or output but interested in coordinates. Examples includes predicting stocks, predict box office collections.

Unsupervised tasks: Training data does not include desired outputs. Example is clustering. It is hard to tell what is good learning and what is not.

Clustering: Clustering images into different categories .No labels associated.

Generation: Based on input, generates similar type values. Only x coordinate exist.

1. **Model:**

Model gives the relationship between coordinates in the form of y = f(x)

Choosing a complex model could often lead to overfitting. We will revisit this when we talk about bias-variance tradeoff.

1. **Loss function :**

Many functions can be approx fit to the model. but how to find to which one is better?.

Using loss functions ,we can determine better approach to see approximation of different models.

Machine can automate the parameters and machine can calculate the parameters for better approach.

Different loss functions;

Squared error loss

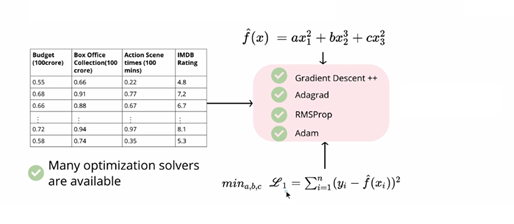
Cross entropy loss

KL divergence

1. **Learning Algorithm**

Colleting of data, task, come up with the function, loss function is the task of human.

Computing parameters in the function is the task of the machine which is part of learning alogorthim .



In the above diagram, input data is in first 3 columns i.e. x1,x2,x3 and output is in the 4th column (IMDB rating) which is y.

The function proposed and loss function shown above.

Machine learning responsibility is to calculate the parameters a,b,c which fit into the function.

1. **Evaluation:**

Evaluation is used to determine accuracy of the function.

Accuracy = Number of correct predictions/Total number of predictions.

Accuracy gives correct precision over loss function.