1.1 Challenges in Machine Learning and Introduction to Probability

1. Inadequate Training Data

Noisy Data-It is responsible for an inaccurate prediction that affects the decision as well as accuracy in classification tasks.

Incorrect data-It is also responsible for faulty programming and results obtained in machine learning models. Hence, incorrect data may affect the accuracy of the results also.

Generalizing of output data-Sometimes, it is also found that generalizing output data becomes complex, which results in comparatively poor future actions.

2. Monitoring and maintenance

Regular monitoring and maintenance become compulsory for the same. Different results for different actions require data change; hence editing of codes as well as resources for monitoring them also become necessary.

3. Lack of skilled resources

Although Machine Learning and Artificial Intelligence are continuously growing in the market, still these industries are fresher in comparison to others.

4. Process Complexity of Machine Learning

Machine learning includes analyzing the data, removing data bias, training data, applying complex mathematical calculations, etc., making the procedure more complicated and quite tedious.

5. Slow implementations and results

Machine learning models are highly efficient in producing accurate results but are time-consuming. Slow programming, excessive requirements' and overloaded data take more time to provide accurate results than expected.

6.Overfitting and Under fitting

Overfitting:

A statistical model is said to be overfitted when the model does not make accurate predictions on testing data. When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set. And when testing with test data results in High variance. Then the model does not categorize the data correctly, because of too many details and noise. The causes of overfitting are the non-parametric and non-linear methods because these types of machine learning algorithms have more freedom in building the model based on the dataset and therefore they can really build unrealistic models. A solution to avoid overfitting is using a linear algorithm if we have linear data or using the parameters like the maximal depth if we are using decision trees.

- Let's understand with a simple example where we have a few training data sets such as 1000 mangoes, 1000 apples, 1000 bananas, and 5000 papayas. Then there is a considerable probability of identification of an apple as papaya because we have a massive amount of biased data in the training data set.

Reasons for Overfitting:

1. High variance and low bias.
2. The model is too complex.
3. The size of the training data.

Techniques to Reduce Overfitting

1. Increase training data.
2. Reduce model complexity.
3. Early stopping during the training phase.
4. Ridge Regularization and Lasso Regularization.
5. Use dropout for neural networks to tackle overfitting.

Underfitting:

• A statistical model or a machine learning algorithm is said to have underfitting when a model is too simple to capture data complexities. It represents the inability of the model to learn the training data effectively result in poor performance both on the training and testing data. In simple terms, an underfit model's are inaccurate, especially when applied to new, unseen examples. It mainly happens when we uses very simple model with overly simplified assumptions. To address underfitting problem of the model, we need to use more complex models, with enhanced feature representation, and less regularization.

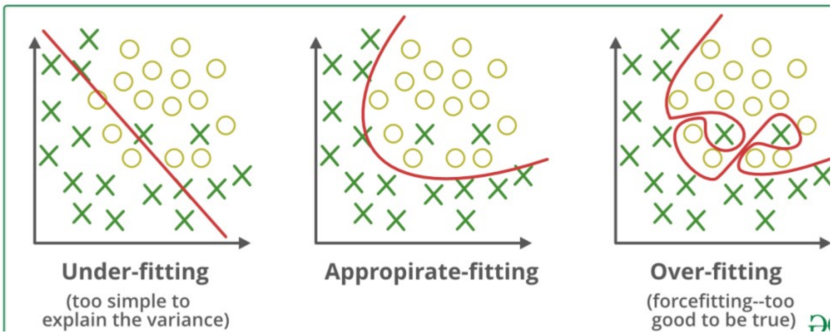
Note: The underfitting model has High bias and low variance.

Reasons for Underfitting

1. The model is too simple, So it may be not capable to represent the complexities in the data.
2. The input features which is used to train the model is not the adequate representations of underlying factors influencing the target variable.
3. The size of the training dataset used is not enough.
4. Excessive regularization are used to prevent the overfitting, which constraint the model to capture the data well.
5. Features are not scaled.

Techniques to Reduce Underfitting

1. Increase model complexity.
2. Increase the number of features, performing feature engineering.
3. Remove noise from the data.
4. Increase the number of epochs or increase the duration of training to get better results.


Under-fitting
(too simple to explain the variance)
Appropirate-fitting
Over-fitting
(forcefitting--too good to be true)

Probability

Probability is **the foundation stone of** ML, which tells how likely is the event to occur. The value of Probability always lies between 0 to 1. 1 indicates more likely that event will occur. 0 indicates that event will not occur.

Formula:

Probability of an event = (Number of way an event can occur) / (Total number of outcomes)

Example 1: Tossing a Coin

When a coin is tossed, there are two possible outcomes: Heads (H) or Tails (T)
Number of ways Head can happen: 1(there is only 1 face with a “H" on Coin)
Total number of outcomes: 2(there are 2 faces altogether)
So the probability of Head (H)= 1/2 i.e 50%, similar for Tail also

Find the chances of rolling a "4" with a single die?

Number of ways it can happen: 1(there is only 1 face with a "4" on it)
Total number of outcomes: 6(there are 6 faces altogether)
So the probability =*1*/**6**

Independent Probability

- Events can be "Independent", meaning each event is **not affected** by any other events.

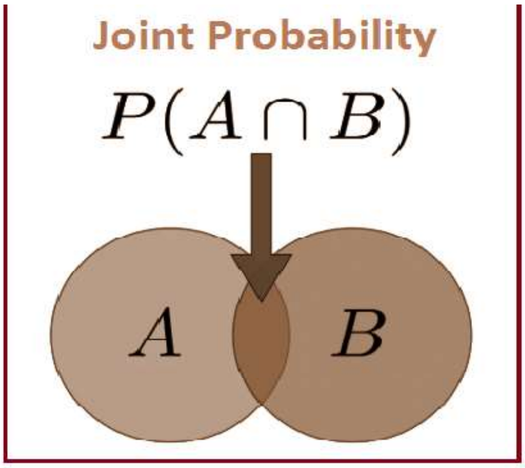
Example:

- Toss a coin three times and it comes up "Heads" each time ... what is the chance that the next toss will also be a "Head"?
- The chance is simply 1/2, or 50%, just like ANY OTHER toss of the coin.
- What it did in the past will not affect the current toss!
- **Also called as marginal probability.**

Joint Probability

- A joint probability is the probability of event A and event B happening at same time.
- The joint probability of event *A* and event *B* is written formally as:
- P(A and B) or P(A ^ B) or P(A, B)
- The joint probability of event *A* and event *B* can be calculated as:

P(A and B) = P(A) * P(B)


Joint Probability
$P(A \cap B)$
A
B