

# NEURAL NETWORKS AND DEEP LEARNING CST 395 CS 5TH SEMESTER HONORS COURSE- Dr Binu V P, 9847390760

CS 5th Semester Honors course for the Computer Science at KTU- Dr Binu V P

## Challenges in Machine Learning



September 02, 2022

Although machine learning is being used in every industry and helps organizations make more informed and data-driven choices that are more effective than classical methodologies, it still has so many problems that cannot be ignored. Here are some common issues in Machine Learning that professionals face to inculcate ML skills and create an application from scratch.

### 1. Inadequate Training Data

The major issue that comes while using machine learning algorithms is the lack of quality as well as quantity of data. Although data plays a vital role in the processing of machine learning algorithms, many data scientists claim that inadequate data, noisy data, and unclean data are extremely exhausting the machine learning algorithms. For example, a simple task requires thousands of sample data, and an advanced task such as speech or image recognition needs millions of sample data examples.

### 2. Poor Quality of Data

Data plays a significant role in the machine learning process. One of the significant issues that machine learning professionals face is the absence of good quality data. Unclean and noisy data can make the whole process extremely exhausting. We don't want our algorithm to make inaccurate or faulty predictions. Hence the quality of data is essential to enhance the output. Therefore, we need to ensure that the process of data preprocessing which includes removing outliers, filtering missing values, and removing unwanted features, is done with the utmost level of perfection.

### 3. Non-representative training data

To make sure our training model is generalized well or not, we have to ensure that sample training data must be representative of new cases that we need to generalize. The training data must cover all cases that are already occurred as well as occurring.

Further, if we are using non-representative training data in the model, it results in less accurate predictions. A machine learning model is said to be ideal if it predicts well for generalized cases and provides accurate decisions. If there is less training data, then there will be a sampling noise in the model, called the non-representative training set. It won't be accurate in predictions. To overcome this, it will be biased against one class or a group.

Hence, we should use representative data in training to protect against being biased and make accurate predictions without any drift.

#### **4. Overfitting and Underfitting**

##### **Overfitting:**

Overfitting is one of the most common issues faced by Machine Learning engineers and data scientists. Whenever a machine learning model is trained with a huge amount of data, it starts capturing noise and inaccurate data into the training data set. It negatively affects the performance of the model. 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; hence prediction got negatively affected. The main reason behind overfitting is using non-linear methods used in machine learning algorithms as they build non-realistic data models. We can overcome overfitting by using linear and parametric algorithms in the machine learning models.

Methods to reduce overfitting:

Increase training data in a dataset.

Reduce model complexity by simplifying the model by selecting one with fewer parameters

Ridge Regularization and Lasso Regularization

Early stopping during the training phase

Reduce the noise

Reduce the number of attributes in training data.

Constraining the model.

##### **Underfitting:**

Underfitting is just the opposite of overfitting. Whenever a machine learning model is trained with fewer amounts of data, and as a result, it provides incomplete and inaccurate data and destroys the accuracy of the machine learning model.

Underfitting occurs when our model is too simple to understand the base structure of the data, just like an undersized pant. This generally happens when we have limited data into the data set, and we try to build a linear model with non-linear data. In such scenarios, the complexity of the model destroys, and rules of the machine learning model become too easy to be applied on this data set, and the model starts doing wrong predictions as well.

Methods to reduce Underfitting:

Increase model complexity

Remove noise from the data

Trained on increased and better features

Reduce the constraints

Increase the number of epochs to get better results.

## **5. Machine Learning is a Complex Process**

The machine learning industry is young and is continuously changing. Rapid hit and trial experiments are being carried on. The process is transforming, and hence there are high chances of error which makes the learning complex. It includes analyzing the data, removing data bias, training data, applying complex mathematical calculations, and a lot more. Hence it is a really complicated process which is another big challenge for Machine learning professionals.

## **6. Slow Implementation**

This is one of the common issues faced by machine learning professionals. The machine learning models are highly efficient in providing accurate results, but it takes a tremendous amount of time. Slow programs, data overload, and excessive requirements usually take a lot of time to provide accurate results. Further, it requires constant monitoring and maintenance to deliver the best output.

## **7. Imperfections in the Algorithm When Data Grows. ( Monitoring and Maintenance)**

So you have found quality data, trained it amazingly, and the predictions are really concise and accurate. Yay, you have learned how to create a machine learning algorithm!! But wait, there is a twist; the model may become useless in the future as data grows. The best model of the present may become inaccurate in the coming Future and require further rearrangement. So you need regular monitoring and maintenance to keep the algorithm working. This is one of the most exhausting issues faced by machine learning professionals.

## Conclusion

An ML system doesn't perform well if the training set is too small or if the data is not generalized, noisy, and corrupted with irrelevant features. We went through some of the basic challenges faced by beginners while practicing machine learning. Machine learning is all set to bring a big bang transformation in technology. It is one of the most rapidly growing technologies used in medical diagnosis, speech recognition, robotic training, product recommendations, video surveillance, and this list goes on. This continuously evolving domain offers immense job satisfaction, excellent opportunities, global exposure, and exorbitant salary. It is high risk and a high return technology. Before starting your machine learning journey, ensure that you carefully examine the challenges mentioned above. To learn this fantastic technology, you need to plan carefully, stay patient, and maximize your efforts. Once you win this battle, you can conquer the Future of work and land your dream job!



## Popular posts from this blog

### NEURAL NETWORKS AND DEEP LEARNING CST 395 CS 5TH SEMESTER HONORS COURSE NOTES - Dr Binu V P, 9847390760

*October 03, 2022*

About Me Syllabus Question Paper Dec 2022 Module 1 ( Basics of Machine Learning)  
Overview of Machine Learning Machine Learning Algorithm Linear Regression Capacity,  
Overfitting and Underfitting Regularization Hyperparameters and Validation

[READ MORE](#)

### Machine Learning Algorithm

*October 01, 2022*

A machine learning algorithm is an algorithm that is able to learn from data. But what do

we mean by learning? Mitchell (1997) provides the definition "A computer program is

[READ MORE](#)

## Syllabus

*October 01, 2022*

Syllabus Module - 1 (Basics of Machine Learning ) Machine Learning basics - Learning algorithms - Supervised, Unsupervised, Reinforcement, overfitting, Underfitting, Hyper parameters and Validation sets, Estimators -Bias and Variance. Challenge: ...

[READ MORE](#)

 [Powered by Blogger](#)

Theme images by [Michael Elkan](#)



**DR.BINU V P-9847390760**

Associate Professor and Head Dept of  
Computer Science-IHRD

Karunagappally.Love to share the  
thoughts and views .I play lot of Games  
and basically an Athlet in my school and  
college days.I never keep my studies as  
a second option.Stood first In MTech  
and BTech.I did my resear

[VISIT PROFILE](#)

**Archive**



[Report Abuse](#)