The Art Of Training Machines

SAI SRAVAN KOTHACHERUVU

The Art Of Training Machines

First Edition

http://www.codemega.com

TABLE OF CONTENTS

PREFACE

Welcome. If you are reading this book, then I assume that you are Interested and curious to know about this fancy term called **MACHINE LEARNING** which is currently the only field of study running the entire world. Right from the time we wake up in the morning, we encounter the use of Machine Learning in many forms. From booking the food for morning breakfast to asking Alexa for setting our morning alarm, we take help of Machine Learning in every walk of our life.

   In recent years, this field has become so advanced that the fancy words coming out of this subject sometimes terrify the students. Students with little or no statistics background should think twice before diving into these immense complicated resources with all the fancy terms that move around the Internet. For all such students who have no knowledge on Machine Learning and want to get a basic understanding , this book is ideal.

   To help such students who want to start their career in Machine Learning and confused on how to start , I am writing this book to help them

understand the basic music that is running in the background. I tried to explain the concept as simple as possible with little or no mathematics involved, to give you a kick start towards Data Science and there by leave it to you to explore yourself. I am also a beginner in this field but I also wanted to share my knowledge that I have gained to all those freshers who wanted to join along.

I want to thank Dr. M Madhu Bala Madam for always supporting me in my projects and also Mrs B. Padmaja Madam for giving me the initial push towards Data Science. I hope this book helps all those students who are confused on how to train machines.

psst. I will not be telling you the pythonic way of doing machine learning because it is available almost every where. I will tell you the way in which the algorithms work. BAM!

Let's Dive In!

This Book is dedicated to my parents and to my inspirational personality, Sir Jeff Bezos

෨෧

INTRODUCTION

Let us assume that Ravi is fighting for world boxing championship and his opponent is Muhammad Ali. I am placing Ravi in a great deal of trouble here! Even though winning the match is less likely for Ravi, he managed to survive the last round of the championship. Now who on earth expects him to survive the fierce punches of the great Ali?

To know how Ravi managed to survive, let us go a little deep and analyse the fight. Ali started fighting by putting his right foot forward, making a furious expression and punching Ravi's face. He then takes a step back and with the same furious expression, stays still without punching.

After that, he takes two steps forward and starts hitting Ravi's face continuously with no expression. Then he takes the same two steps forward with smiling face and punches Ravi's face.He then Stands still and roars heavily without punching Ravi. Then Ali turns around and starts laughing and asking audience to cheer for him.

Initially Ravi had to suffer a few punches but after a while he managed to defend Ali. Forget about Ravi. Here, Even you can defend Ali's

punches! By observing Ali's fight, we can draw some insights as follows:

BODY MOVEMENTS	EXPRESSION	ACTION
Right foot forward	Furious expression	Punch face
Takes a step back	Furious expression	Doesn't punch
Two steps forward	No expression	punch face
Two steps forward	smile	punch face
Stands still	Roar	Doesn't punch
Turns around	smile	Doesn't punch

Enough is enough! Ravi suffered a lot of punches. Now it's time to defend.

After asking audience to cheer, Ali comes back and

Case 1. Takes his right foot forward.

Case 2. Takes two steps forward.

Case 3. Stands still.

Now can you say if Ali punches or not for these above cases?

If you can say so, then congratulations! Even you can survive till last round.

Now let us try to find the actions for above cases.

In case 1, he took his right foot forward. From the above table, regardless of the expression, we can say that if the body movement is to take right foot forward, then Ali is going to punch. In this

way, Ravi predicted the action and defended Ali's punch.

BODY MOVEMENTS	EXPRESSION	ACTION
Right foot forward	Furious expression	Punch face

For case 2, From the table, even though he smiled as well as shown no expression for same body movement of taking two steps forward, he punched in both cases. so be it any expression, if he takes two steps forward, then he is going to punch!

BODY MOVEMENTS	EXPRESSION	ACTION
Two steps forward	No expression	Punch face
Two steps forward	smile	Punch face

Now I will leave the case 3 for you.

So in this way, Ravi was able to analyse the previous moves of Ali and predicted his upcoming actions based on previous moves and expressions.

In machine language lingo, these body movements and expressions are called **FEATURES** and Actions are called **LABELS.**

Here Ravi has taken help of previous body movements,expressions and actions to analyse and identify patterns and finally predicted the actions for the new body moments and expressions. So Ravi learnt from that data, analysed and

understood. Then he predicted the actions of Ali based on his new body movements and expressions. Now we can call this as "Ravi Learning" because he is getting trained by the data that he collected and then was able to attend the test and predict the upcoming actions based on his training efficiency. BAM!

Note: In Machine Learning lingo, the data used by Ravi for learning and training himself is called **Training Data** and the data that he encountered afterwards (movements by Ali after coming back in to the fight by asking people to cheer) to predict is called **Unseen Data.**

Now if we can use the same data to train a Machine(computer) instead of Ravi to predict the actions, the process is called "**MACHINE LEARNING**" and the algorithm(Instructions which guides machine in learning and training) used to train the machine is called machine learning algorithm. DOUBLE BAM! There are many machine learning algorithms based on the type of data we use and also based on the type of results that we want. Machine Learning can generally be classified into three types:
  1. Supervised learning
  2.Unsupervised learning
  3.Reinforcement Learning

Simply put, in **Supervised learning,** a machine is provided with some initial features along with labels, so that it gets trained and prepares itself for the testing data to predict the labels for the new unseen features. It's just like Ravi taking help of the initial actions corresponding to body movements and expressions to train himself to predict the upcoming actions of Ali. The outcomes in this learning would be specific.

In **Unsupervised learning,** there are no outputs or labels provided and the machine will train itself by analyzing the patterns and structure of the features given. It is Body Movements and expressions without Actions, if you want to consider the above example itself! There is no specific output for unsupervised learning and it helps in finding hidden patterns in the data.

I don't want to discuss much about Reinforcement learning but it generally trains machine on how to act in a specific environment.

If you didn't really catch the meaning of these different machine learning types, don't worry. It will be clear to you as we proceed further.

TYPES OF DATA:

Have you observed the data collected by Ravi to analyse Ali's fight? For every different move that Ali took, Ravi had collected features and also the actions, that belong to the following set:
{ 'punch

face', 'Doesn't punch'} . Do we really call this data? If yes, then what about the whole text that I typed to write this book? What about all those giant looking numbers, text messages, emails etc. which gives us a lot more information than this mere mathematical set containing two strings? Well, make yourself clear that **Data is some information and it's form and structure varies based on the context and situations.** Data can be in any form. From numbers to words and giant paragraphs, From pictures to graphs, Data can take on any form and structure.

In the field of machine learning, we encounter different types of data which we need to handle to derive an efficient Machine learning model that could help in prediction and decision making. In the above example, Ravi handled **Categorical data** i.e., either punch or doesn't punch. This type of data specifies to which category, a particular data entry belongs.

Ex: (Male,Female),(Yes,No),(Rainy,Sunny,cold) etc

Categorical data can be represented in the form of numbers ('punch'=0, 'Doesn't punch'=1) but these numbers doesn't have any mathematcal significance. They are simply used in Machine learning algorithms to ease computation! small BAM!

The other type of data that people generally encounter is **Numerical Data.** This type of data includes numbers which are continuous like 80,80.3,81.789,91.324 etc. and discrete like 30,40,50,60,90 etc. Have you observed the difference between continuous and discrete numerical data? There are other types of data which also have there importance but these two data types are most frequently encountered and used. Medium BAM!

That is enough! We have learnt the types of data that we can encounter while we work with ML models.

You shall eventually understand the impotance of different types of data in formulating a specific Ml model. As we proceed, you can see different Ml algorithms being discussed. I encourage you to relate those algorithms and the type of data used in it with the types discussed here. BAM!

DATA SCIENCE :

Data Scientists job is currently one of the most fascinating jobs around the globe. But what do they actually do? What is Data Science? Why are we even studying about it?

I will make things easy for you here.

Data science is about using data and creating as much impact as possible on your company.

Impact here can be in any form. It can be insights, product recommendation etc. It is a field of study which is used to solve real world problems using data. BAM!

Let us now understand the history of data science to get more insights on the field.

Data science has never taken it's origin directly in a straight forward way. In the year 1996, an article called "FROM DATA MINING TO KNOWLEDGE DISCOVERY IN DATABASES" discussed about a concept called **Data Mining**. In this article, the process of extracting useful information from data using different algorithms is called as Data Mining. It is referred to finding patterns from a huge set of data for understanding data and using it in different real world applications. In 2001, William S.Cleveland took Data Mining to a completely different level. He took computer science and combined it with Data Mining to form Data Science. Double BAM!

Data Mining + Computer Science is equal to Data Science

William S Cleveland

Around this time, WEB 2.0 also emerged where websites were no longer digital templates but a medium for shared experiences among millions of users. Myspace in 2003 and Facebook in 2004, to name a few were among those sites with millions of traffic. People can then post a picture,like a picture, comment a picture, share a picture,chat with their friends and connect to all the people and events happening around the globe. All these activities gave rise to a new type of data called **Big Data.** With millions and millions of people interacting and using these sites, the data that is generated is immense! To handle such a huge amount of data efficiently, we need different tools and technologies. Unlike traditional techniques to store and process the data, we need software that could perform parallel processing and storing this data which is big, in an efficient manner. Ravi did managed to collect

the data and analyse it. But think about millions of people like Ali with different fighting strategies. Can Ravi still be able to collect and analyse such huge data? Even a normal computer cannot handle such immense computational tasks. That is the idea of **Big Data!** BAM! To handle big data,software utilities like HADOOP and SPARK were created. I won't go deep in to this topic but I encourage you to get your hands dirty by installing and playing around with these software applications to get a clear idea on the things that are running behind handling Big Data. Generally these software utilities split the computational task and instead of giving the whole burden to a single system, it spreads the work load to different systems such that the speed of computation would be fast and easy. Ofcourse, using more labor to build a house is sensible than using a single labor.

That is enough! Are we loosing our track from Machine Learning towards Data Science and Big Data?

Let me explain you the steps involved in Data Science. Then you can answer the above question yourself.

Data science involves tasks like:

1. Data collection, Data organisation and Data cleaning, where Data Engineers collect the data from different sources, store them efficiently