Introduction To ML

Machine learning

Noel Moses Mwadende
TANZANIA MoTech

ABOUT THE AUTHOR

Noel Moses Mwadende is currently 2 year Computer Science and Information Security student at University of Dodoma, Data Scientist and well specialized inCyberSecurity.

Apart from this books other books written by same author , which are present at Github https://github.com/MoTechStore are "WIFI HACKING IN FEWS STEPS" ,"PANDAS TOOL FOR DATA SCIENTIST",”ZIP PASSWORD CRACKER” and “INTRODUCTION TO MACHINE LEARNING”.

Currently the author has been employed at MoTech YouTube channel (https://www.youtube.com/channel/UCtuaigKZF3okQnKON5RM1qQ/playlists), and so far the author have produced more than 70 tutorials about python, machine learning, pandas, VB.net, php and other Information Technology.

INTRODUCTION

This book covers basic concepts for beginners in machine learning, explaining different types of machine learning, algorithms consisted in each type of machine learning, and how to do machine learning.

Inshort the intention of this edition just to make awareness and derive this new concept, because it happens in most cases, some beginners of machine, just start with learning algorithms, while ignoring the basic concepts which are basic for their field.

ACKNOWELDGEMENT

I would like thanks to my instructors of machine learning and Data Science Sir Salim Diwani, Mr Eliah and Miss Martha Shaka, to be honest these are the most import people in my , because I can not am good data scientist without mention their names, they made forgotten.

Thanks to my security instructor Sir L Mutembei, your role in my development in writing books, producing tutorials at MoTech, and of machine learning for cybersecurity.

Thanks to MoTech for sponsoring of the writing of this book to be host without them for sure I could prove some failure in writing this book.

Thanks to Kelvin N Simchimba currently MTA 2 year student at UDOM, though I contacted you so late but you made good design of cover page and the whole structure of the book.

Lastly, I would like to thanks my friend JohnBosco Francis because my first lecture to you about machine learning introduction was the starting point of this book, because after that session then I thought why can't I teach others about machine learning, how ?, it is through this book was made.

Table of Contents

INTRODUCTION OF MACHINE LEARNING

Machine Learning is the system that can learn from examples through self-improvement and without explicitly coded by programmer, the meaning come with idea that machine can learn from data or examples to produce accurate results. Machine learning is closely to data mining and Bayesian predictive modelling.

Simple example where machine learning can function as human being, consider this scenario, for old man lived 85 year , coming across through different summer and winter seasons can predict the occurrence or thee rate of rainfall in the next year, how this becomes possible ? it is due to experience, also for machine learning given a Dataset about rainfall statics organized in good way, also computer can learn from those Dataset and come up with prediction.

A football fan haven watching Manchester United Vs Liverpool games for 10 years, have data, experience, and can predict the game who is going to win the match, this is same for machine learning, by using different machine learning algorithms and Dataset, the system can be made possible to predict.

Arthur Lee Samuel (1959) described machine learning as "It is the field of the study that gives the ability to the computer for self-learn without being explicitly programmed ". machine learning, learn itself from Dataset exposed to them and finally they come with great output.

It good view from Arthur Lee Samuel, but if machine learning project is deployed by yourself it will need programming language concepts, since a lot of people nowadays are using PHP, Django and Flask for deploying their projects.

COMMONLY MIXED TERMS IN MACHINE LEARNING

Data science

is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is very wide field as it covers a lot of things, for someone to be specialist in data science will need a lot of combination knowledge like programming language, idea about mathematics and statistics, database, it also involve machine learning.

Machine Learning

is the field of computer science that deals with giving ability to the computer to self-learn and do a work which may be prediction or any other intended work .It will need Dataset which will be exposed to the algorithm, for it to learn, in supervised machine learning Dataset are usually labelled, algorithms learning from features containing names, but this is quite different to unsupervised machine learning where Dataset have no labels, so the algorithms learn independently.

Big Data

is the field that deals with the ways to analyse, systematically extract information from data set that are too large or complex to

deal with by the traditional data processing application software. Data with many rows offer greater statistical power, while data with higher columns may lead to higher false discovery rate.

Data Mining

is the process of discovering patterns in large data sets involving methods such as machine learning, statics and database systems.it is an interdisciplinary subfield of computer science and statics with an overall goal to extract information from data set and transform the information into a comprehensible structure for further use .Different algorithms are used to extract data from social media, which in turn is used for health centres and production of products according to the desire of customers.

Algorithms

are the methods or procedures used to get the task done or to solve a certain problem sometimes these algorithms are called classifiers.
These algorithms are already developed, the remaining task is to import and use them. Sklearn is library which contains machine learning algorithm.

Models

are well defined computations formed as the result of an algorithm that takes some values as input and produces some value

as output .so model is the result of algorithm, they take some values as input, process them and come with output.

RELATION BETWEEN DATA SCINCE AND MACHINE LEARNING.

Data science

is wide term, covering many fields like machine, statics, analysis and some concepts of programming language. machine learning is direct term which involve algorithms, to be data scientist is not necessary to have knowledge of machine learning, but also for good data scientist will need you to cover machine learning.

Machine learning

includes some techniques that can be useful for data scientist, but data science does not more rely on machine learning, but data science is much more broad or wide.

TYPES OF MACHINE LEARNING

1. **Supervised**
2. **Unsupervised**
3. **Semi-supervised**
4. **Reinforcement**

1. SUPERVISED MACHINE LEARNING

is the type of machine learning that learn from past input data and make future prediction as output.is the machine learning in which labelled data used to train the algorithms, algorithm are trained using data with labelled names where the data is divided into train and test same like input data and output data, these input data are called features, indicated by x and output data which is indicated by y.

SUPERVISED MACHINE LEARNIG METHODS

A. **Classification (yes or no, High or low)**

B. **Regression**

A. **= = => Classification (yes or no, High or low)**

This is concerned with building models that separate data into distinct classes, it always deals with problem which have binary output, that is two output.

Classification Algorithms

= Logistic Regression

= K-Nearest Neighbours

= Support Vector Machine

= Kernel Support Machine

= Naive Bayes

= Decision Tree

= Random Forest

B. **= = => Regression**

This is based on taking input data and then machine learning predict continuous output values, as you can see that the data differs between classification and regression learning, in classification the output is binary but in regression learning the output is continuous.

Classification Algorithms

= Simple Linear Regression

= Decision Tree

= Random Forest

= Support Vector Machine

SEMI-SUPERVISED MACHINE LEARNING

It uses small amount of labelled data and large amount of unlabelled data this is contrary to supervised machine learning which uses complete labelled data. It is mostly used for speech recognition and classification of texts.

Semi-supervised Algorithms

= Graph Based Algorithm

= Generative Models

= Self Training

REIFORCEMENT MACHINE LEARNING

Is this machine learning, model are able to learn based on reward and punishment receive for it's previous actions, it make decision sequentially according input and out to the model.

Reinforcement Machine learning Algorithms

= SARSA

= DDPG

= Deep Q Network

= Q- Learning

2. UNSUPERVISED MACHINE LEARNING

is the type of machine learning that try to find hidden structure pattern by using unlabelled data, the model is given data or examples then after understanding the new data is given to the model to test it.

UNSUPERVISED MACHINE LEARNING METHODS

A. Clustering

B. **Association**

A. Clustering

This is used for analysing and grouping data which are not labelled class or even a class attribute at all.it is algorithm which deals with grouping sets which are similar into clusters.

Clustering Algorithms

= K-Means

= Hierarchical clustering

= Hidden Market model

= Fuzzy C-Means

B. = = => Association

Discover the probability of co-occurrence, how can multiple item occur together, how the occurrence of one item have an association to the occurrence of another. Most employed for recommendation systems and arranging of products in supermarket.

For example, pen, pencil and exercise book can be associated together, mil and bread can be associated or grouped together.

Association Algorithms

= Aprior

= Eclat

STAGES OF MACHINE LEARNING PROJECTS

1. **Data Collection.**
2. **Data Preparation.**
3. **Choose a model.**
4. **Train the model.**
5. **Evaluate the Model.**
6. **Parameter Turning.**
7. **Make Predictions.**

1. **Data Collection.**

This step involving getting or finding data which will be used for your project, one can get data direct from organization, but if it is normal research consider your selected area of research ,for beginner to machine learning there are many repository UCI ,Kaggle ,Google Dataset research and visual data containing a huge amount of data.

2. Data Preparation.

This step will involve cleaning of data, and changing them to numeric type, so if data contain question mark, quotes, punctuation should be removed, removing of extra white space, removing duplicated data, and making all missing values are removed by either filling them or totally removing them.

3. Choose a model.

Criteria for choosing a model is according to the problem you solve, for example if machine learning is concerned with output which is binary , regression algorithms like logistic regression will be used, but if machine learning is about predicting continuous values like price linear regression will be used, but sometimes the prediction may be multiple where one can choose to use random forest.

4.Train the model.

Divide Dataset into train and test, almost train and test data are in ration of 0.7:0.3, most of the time training data is 70% and testing data is 30%, then after data is divided, fit the model.

5. **Evaluate the Model.**

Checking the result of model, cross validation is used evaluate performance of the model, where performance of the model can be cross validated according to the value of k (k fold).

8. Parameter Turning.

These are made internally in algorithm, that are required when using a model, for example in linear regression and logistic regression coefficient is their parameter and support vector in support vector machine, so these parameters usually need to be passed in algorithms for it to function well.

According to different algorithms there are parameters which are internal configure in the model, but there is hyperparameter which are external to the model, this part is about optimizing a model.

9. Make Predictions.

After everything is set clear use it for making prediction.

HOW MACHINE LEARNING WORKS

It is like the way machine learn is like human being, human learn from experience, as long as we know is when our ability to handle different situation grow, it is where we can predict different things in the society, for example an old man who has lived about 125 years , passed through many summer and winter season in his lifetime can easily predict when will the rain start for the next year due to his experiences, it is the same way machine learning works , given many data composes of many rows and columns about rain statics of the past year it will become easy for it to predict when or at what amount the rain will be.

APPLICATION OF MACHINE LEARNING

1. **Search engine result**
2. **Voice recognition**
3. **Face recognition**
4. **Prediction**
5. **Spam filtering**
6. **Recommendation system**
7. **Social media**
8. **Sentiment analysis**
9. **Video surveillance**
10. **Speech recognition**

CONCLUSION

Thank you that was the end, In next edition will be about supervised machine learning algorithms, this was just introduction, it is better you understand all basic, so that when comes an issue of choosing algorithm, you should know when and where to use the model, and why.

Digging deep into this book will turn you into one of Best in machine learning, for sure it will move you to a certain level, for those who takes data science, python and machine learning https://www.youtube.com/channel/UCtuaigKZF3okQnKON5RM1qQ/playlists), is the link you can get my self made tutorial about python, pandas and machine learning, but also https://github.com/MoTechStore is link to get my previous books, and will keep on adding books , welcome and review.

REFERENCES

1. Duda, R.,Hart and Stock,D.(2001). The Elements Of Statistical Learning – Data Mining, Inference and Prediction.Berlin:Springer-Verlag.
2. Mitchell,T.(1997).Machine Learning. New York: Mc Graw-Hill.
3. Bishop,C.M.(1995). Neural Networks For Pattern Recognition. New York: Oxford University Press.
4. Baldin,P. and Brunak,S.(2002). Biometrics: A Machine Learning Approach, MA: MIT Press.
5. Goodfellow,I.,Bengio,Y and Courville,A.(2016). Deep Learning. MIT Press
6. Christianini, N and Shawe-Taylor,J.(2000). An Introduction To Support Vector Machine. London: Cambridge University Press.
7. Vapnik,V .(2013).The Nature Of Statistical Theory. Springer Science & Business Media.
8. Jollifie, L. (2002).Pricipal Concept Analysis.Wiley Oline Library.
9. Tan, P-N., Steinbach., and Kumar, V.(2004). Introduction Ta Data Mining. New York Addison-Lesley.
10. J.O.Berger,.(1985). Statistical Decision theory and Bayesian Analysis,Springer.New York
11. C.R.,Rao.(1973). Linear Statistics inference and it's application,John Wiley and Sons.New York
12. G.R. Shorack and J.A Wellner. (1986).Emeprical Process with application to Statistics,Wigley.New York