Acceleration of Machine Learning Algorithms

NISHANT CHOUDHARY

(**TEQIP_INTERN**)

ABOUTS:

EDUCATION:

Nishant Choudhary is an undergraduate in Computer Science and Engineering at Atal Bihari Vajpayee Govt. Institute of Engineering and Technology Pragatinagar, Shimla (H.P.), India

Working under Faculty Supervisor:

Dr. Sparsh Mittal

Assistant Professor, Department of Computer Science and Engineering, IIT Hyderabad

Research Interests:

Processor architectures for machine learning, neural network accelerators, computer architecture (CPUs and GPUs), VLSI, high-performance computing, approximate computing.

Guided by:

Mr. Abhijith Girin N V

MTech in Computer Science and Engineering, IIT Hyderabad

Internship Provided by:

Technical Education Quality Improvement Programme (TEQIP) Cell, IIT Hyderabad.

Internship Duration:

One Month (11th June'19 - 10th July'19)

Abstract:

Machine learning is an application of Artificial Intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.

Machine learning algorithms are often categorized as supervised and unsupervised. Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled.

Acknowledgments:

My contribution:

We have found out that various machine learning algorithms have a varying effect on the accuracy of the model. Before implementing any machine learning algorithm the given training data must be analyzed. The first and foremost task that the data analysts do is to view the data and tries to make some sense out of it. This is where EDA(Exploratory Data Analysis) comes into the picture. Here we used techniques such as data statistics, data quality, data visualization, data relationship, dimensionality reduction to get conclusions out of data. After analyzing data we used various machine learning algorithms such as decision tree, SVM and calculated the accuracy of the model with and without using boosting techniques such as catboost, xgboost.

We used EDA and catboost classifier and calculated accuracy on Titanic dataset. We, later on, exchanged catboost classifier with xgboost and calculated accuracy which increased by some percentage.

Conclusion:

For solving the major problem of Machine Learning models, the choice of machine learning algorithm does matter. There are other factors that can also play a major role in increasing the speed of computation and the accuracy of the model like EDA, Feature Engineering, Ensemble methods(Boosting) and Cross-Validation.

Learning Accomplishments:

I started my internship with not having any background from the Machine Learning and from that day to now I have learned a lot. First of all, I was not familiar with Linux OS but now I am quite familiar with Linux file system, administration, automation,scheduling, open source softwares,. I also learned Python programming language and used it to make a game named TIC-TAC-TOE. I learned various libraries used for machine learning such as numpy, pandas, matplotlib, seaborn, sklearn. For doing all these things, my guide Mr.Abhijith Girin N V had guided me.He gave me the right way to learn by providing study material and tutorials. I have also done hands-on experience with numpy, pandas,matplotlib and an overview of CNN which I learned from the teaching session taken by Mr.Ritesh Gupta and Mr. Sharath R . I have learned writing research and survey papers from seminars taken by Dr.Sparsh Mittal. He also gave the knowledge of Latex for writing such papers. I learned how to carry EDA on the dataset, perform training, testing, and prediction using various machine learning algorithms. I also joined Kaggle contest on Titanic Survivors prediction under the guidance of my guide Mr.Abhijith Girin N V.