Mathematics In AI

Rebecca Huo

Exeter Mathematics School

Abstract

This project is based on the investigation of Mathematics used in Artificial Intelligence with a special focus on Calculus and Linear Algebra. For more information on Statistics used in AI, please see my report.

Introduction

This project is a research based project and so there aren't necessarily any final results involved. My approach to start this project by outlining the sections of Maths I want to cover and there are a total of three main areas including: Statistics, Calculus and Linear Algebra. In the report all three sections will be covered and in Statistics, the report will cover mainly Linear Regression and there is a topic called Gradient Descent which can technically be classified under both Statistics and Calculus, but I have decided to cover it in Calculus as it would be easier to comprehend since it involves gradient. In Linear Algebra, there is a strong focus on Graph Theory and and Spectral Clustering because the combination of them can achieve Dimensionality Reduction.

1 Basics of AI

AI, also known as Artificial Intelligence is a major part of the developing technology and so far it has served many purpose from making predictions to entertainment. The most noticeable would the AI chess and go algorithms which managed to beat some of the best human players with just a few days of training. Within AI, there are many different branches and within those, and some of the significant branches include Machine Learning and Deep Learning. Both depends on large data set to train the algorithm, Machine Learning leans more towards prediction whereas Deep Learning leans more towards optimisation. As new as the idea of computer intelligence might be, there has been such ideas since ancient Greek time and

2 Calculus and statistics

2.1 Calculus

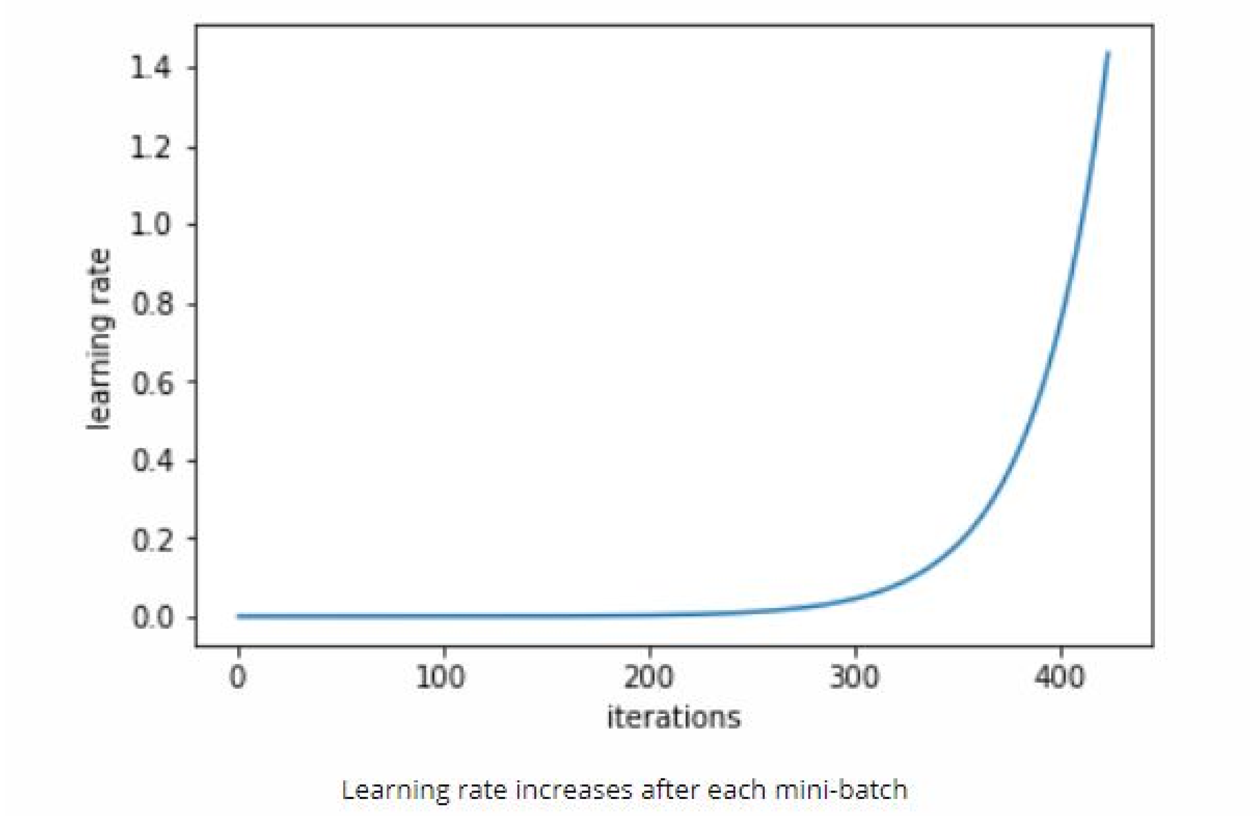
Gradient or Rate of Change is the very fundamental part of Calculus and in AI, it can be used for optimisation which is essentially finding the best solution to solve a particular problem. This is where Gradient Descent is used, first a plot the cost function and then by defining a Learning Rate Parameter, which is the fixed length of a step one can take on the curve to reach the minimum and that is the point where the error between the actual value from the data set and the prediction made by the algorithm the least. With this, we can plot the Linear Regression line in order to find the optimal line of best fit to make predictions. There are currently three types of Gradient Descents:

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini-Batch Gradient Descent

Batch Gradient Descent is where the the algorithm loops through all data points and sums all the costs and then finds the minimum with one iteration of the data set. On the other hand Stochastic Gradient Descent requires the data set to be well shuffled so it can pick data randomly, and then instead of loops through all the data to find the minimum, it finds a new point on the graph after looping through each data point. Mini-Batch Gradient Descent is like the combination of the two, it divides a large data set into smaller sets and make changes on the graph after iterating each mini-batch of data. All three types of Gradient Descents has a common disadvantage and that is if the data set is way too large, it will take a very long to time to find the

minimum point.

Mini-Batch Gradient Descent is fairly new compare to the other two but it's used more and more by people as it provides an exponential growth of Learning Rate against iteration.



[13]

2.2 Statistics

It is all fairly easy to understand why statistics is needed in AI, as AI algorithms require a large data set to learn in order to complete their task, and statistics is needed to handle the large data set. A vital component in machine learning is Linear Regress. The Regression line is commonly used to make predictions are optimisation. To find the regression line, one needs to first calculate the correlation coefficient and that is essentially the gradient of the line. As we are aware that Gradient Descent is needed to form the Regression line and to form the regression line, it will need the weight. A weight in this case is basically the coefficient. In order to obtain a new value for the weight, it needs the previous value for the weight and then subtract by the product of the learning rate and error. Error is gradient and learning rate is as we know a pre-set value.

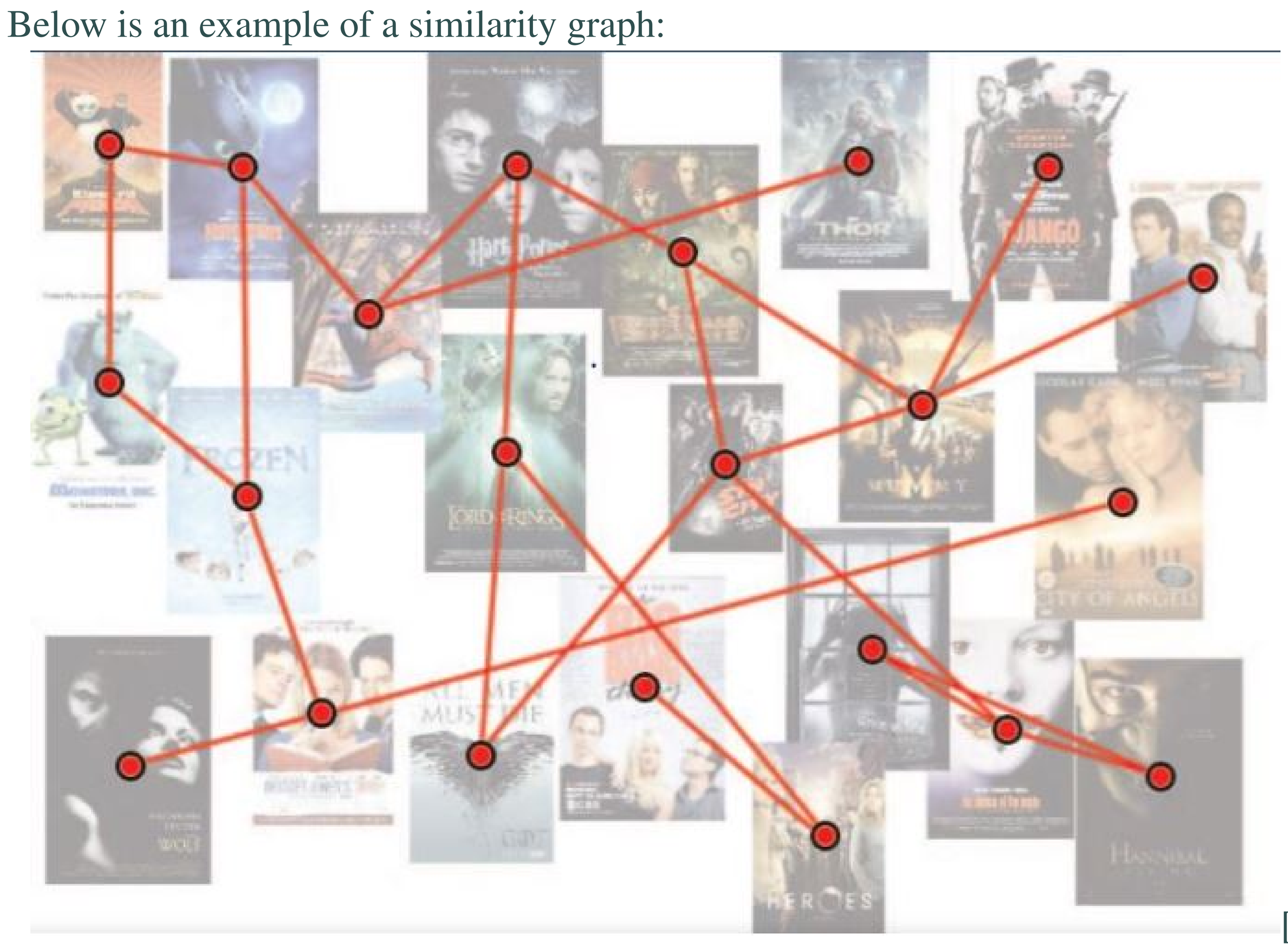
Linear Algebra

Linear Algebra plays a vital role when visualising vector space and so it has many use in computer vision and many other uses. For instance, with the help of Graph Theory and Spectral Clustering, an algorithm can perform Dimenionality Reduction.

Graph Theory is where the relationships between objects can be displayed on a graph. The Graphs can be considered as an upgraded tree where the flow can be both direction. This way, one can create a similarity graph where the algorithm can easily spot the similarities between the objects. Spectral Clustering uses the similarity graph to perform Dimensionlaity Reduction in the following step:

- Create a similarity graph between the N objects that needs to be clustered
- Compute the first K-eigenvectors of its Laplacian matrix to define a feature
- Run K-means on the features to separate the objects into K-classes

Below is an example of a similarity graph:



[6]

As for Dimensionality Reduction, an easier example is to reduce a 3-D object into a 2-D object, then into a 1-D object. This is useful in Machine Learning is sometimes there are too many factors and it would be difficult for the algorithm to process and so with Dimensinality Reduction, it merge profoundly similar features into one. A real-life exmaple of where Dimensinality Reduction is used is in classifying emails and so the the email algorithm can automatically identify spam emails and prevent it from showing up in the inbox

Conclusions

The purpose of this project is to investigate the mathematics used in Artificial Intelligence, however, given the volume of knowledge, I have picked the most essential parts and we have investigated the following aspects:

- Gradient Descent
- Linear Regression
- Graph Theory
- Spectral Clustering
- Dimensionality Reduction

References

1. https://towardsdatascience.com/spectral-clustering-for-beginners-d08b7d25b4d8
2. https://www.kdnuggets.com/2017/11/estimating-optimal-learning-rate-deep-neural-network.html
3. https://machinelearningmastery.com/linear-regression-tutorial-using-gradient-descent-for-machine-learning/

Acknowledgements

I would like to thank all EMS stuff and all those for the help and support they have given me during this project.