Prakhar Saxena

Professor Edward Kim

CS 383 - 001

May 14, 2020

Honours Project: L1 Normalisation

In machine learning, one of the critical tasks is selecting important features that are to be used in a Machine Learning model. These features are often hand selected by the developer; however, their relative importance to the task at hand is unknown. In this project, I use a mathematical regularisation technique, called L1 Normalisation, to identify the features that are most important to a machine learning model.

Overfitting

“The production of an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fir additional data or predict future observations reliably.”

“An overfitted model is a statistical model that contains more parameters than can be justified by the data.”

Regularisation

In the simplest terms Regularisation is a process of adding information to solve overfitting in a Machine Learning Model. It artificially discourages complex equations and solutions, even if they fit perfectly with the training/observed data. Explanations or solutions typically of those complexities do not bode well in generalising a real-world test data, because it fits too closely to the training data leaving any space of freedom for the test data.

References

<https://www.lexico.com/definition/overfitting>

Everitt B.S., Skrondal A. (2010), *Cambridge Dictionary of Statistics*

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html>