

Clustering Algorithms for Color Segmentation

Roger Pujol

Universitat Politècnica de Catalunya (UPC)

Barcelona, Spain

roger.pujol.torramorell@est.fib.upc.edu

ABSTRACT

General Matrix Multiplications (GEMM) are a key component in the rising world of Deep Learning. Most of the computational complexity of Deep Neural Networks can be reduced to multiple huge GEMMs. Since any small optimization in GEMM can affect positively in many applications, our objective in this project will be to find how the parameters affect the performance of this kind of operations. Also we will try to get a regressor model capable to predict the time that will be needed to compute the GEMM given the parameters used.

The code of this project is Open Source and can be found in: https://github.com/rogerpt32/adm_paper2

6 DECISION TREE REGRESSOR MODEL

6.1 Implementation

7 CONCLUSIONS

1 INTRODUCTION

2 DATA

3 PREPROCESS

4 ANALYSIS

5 REGRESSION LINE MODEL

The first approach that we used to try to predict the execution time of the SGEMM given the parameters, is a simple regression line.

5.1 Implementation

To get the model, first we split the model in two blocks: one for training (in this case 90% of the data set) and the other to test (in this case 10% of the data set) the accuracy later. Then we use the training part to fit the regression line model. Finally we use that model to predict the execution times of the test examples and we compare the results with the real values.

5.2 Results

The results from the Regression Line are not very satisfying. The mean square error that this model has in the training data is 80322.28, and obviously the mean square error in the testing data is even worse with a value of 82841.39. This lack of accuracy using a linear regression, might probably mean that our data is not linear. If we check the variance using the R^2 -score, where a 1 means that the data fits perfectly to the regression line with no variance and a 0 means the opposite, we get a 0.41 in both train and test data. This confirms that this regression might not be a good solution.

But in order to benefit from this model, we can take a look to the coefficients that define the model. These coefficients multiply the correspondent input variables described in the Data Set section.