Data-based Modelling Using Machine Learning

UPPSALA UNIVERSITET

Students
Kettler, Kristian Morten
Lundstedt, Håkan
Rivera Öman, Marcus

Supervisor
Mishchenko, Kateryna

Course coordinator
Neytcheva, Maya

Project in Scientific Computing, 2017

Department of Information Technology

Background

Machine Learning has been getting a lot of attention for Big Data problems, such as user data or image recognition. The objective of this project was to examine Machine Learning on the cloud platform (Microsoft Azure Machine Learning Studio) for an engineering problem. The investigation includes comparison of different algorithms and their suitability on the given data points.

Algorithm Selection

Suitable for Problem?

☒ **Linear regression** are methods to approximate dependent variable values by linear functions, most fundamentally by using a least-squares approach.

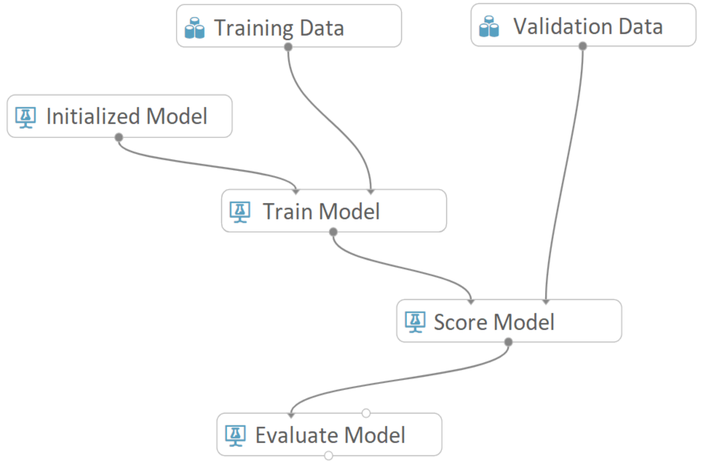
☑ **Artificial Neural Networks**:
A simulation of biological neurons can be taught to solve complex problems and is especially used in recognition problems.

☑ **Boosted Decision Tree:**
Decision Trees determine outputs according to set up rules. Setting up these rules most efficiently can be computationally complex.

Microsoft Azure

Software & Methodology

Azure Machine Learning Studio is a “drag ’n’ drop” solution for quick model set ups. Data can be uploaded to the cloud environment and pre-processed within a model. The user can choose between different initial methods and model parameters. Model training and scoring are automatized processes that give out both results and performance statistics. Model initialization requires understanding of the underlying algorithms.



Schematic model layout in Microsoft ML Studio.

The problem that ABB presented did not have sufficient data for proper method analysis. Therefore artificial measurement data was used to scale the problem from 50 to 200.000 data points. As performance indicators, execution time and accuracy were chosen. A good model is not slowed down too heavily by big data and has a high accuracy.

Conclusions

Azure Machine Learning Studio offers a simple method to implement machine learning on your datasets. In conclusion it is more questionable if machine learning is a suitable for the given problem. To use machine learning as a suitable alternative to analytical models, the number of data points has to be sufficient for both training and validation. With the given number of data points in this project, machine learning can be completely neglected as an alternative. Testing of algorithms has been conducted with artifical data generated from an analytical model and in theory machine learning seems worthwhile.

Outlook

Should ABB decide to use machine learning for modelling in future problems, they have to entsure a sufficiently high number of data points is collected. In automatized experiments, so-called stream analytics can be implemented to set up automatized machine learning processes in Azure (using “re-training”).

Although Microsoft Azure Machine Learning Studio offers a quick-start into machine learning, other software solutions might be more feasible due to broader algorithm choices and efficiency.

It is important to note that Machine Learning Studio is under ongoing development and fairly new.

ABB