# Bootstrapped Learning of Machine Learning: A Case Study using Microsoft Azure ML

Harry Strange
Department of Computer Science,
University College London,
London, WC1 6EA, UK
h.strange@ucl.ac.uk

**Abstract:** This paper introduces a module being taught at University College London that seeks to engage students from a range of engineering backgrounds with the topic of machine learning. Microsoft's Azure Machine Learning platform is used to enable students to build and test machine learning models from the very start of the course. The pedagogical framework behind the module is discussed along with some practical insights into how Azure ML is used. Conclusions show that this approach could be deployed to a wide range of students from different academic backgrounds.

## INTRODUCTION

> *"If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth ten Microsofts."*
> - Bill Gates, (Lohr, 2004)

There is little doubt that we are experiencing an unprecedented increase in the volume, variety, and velocity of data created every day. This "data big bang" (Penson et al., 2010) has led to the creation and collection of vast quantities of data; however, only a fraction of this data – estimated at less than 0.5% – is every analysed and used (Marr, 2015). The reasons for businesses leaving this data unanalysed are myriad and complex. One of the foremost reasons for this is the gap in available expertise, a fact evidenced by the high starting salary that many data scientists are offered (KDnuggets, 2014).

It has been estimated that businesses will need one million data scientists and machine learning practitioners within the next few years (Forrest et al., 2016). Although many businesses at present rely on in-house training to develop the skills of data scientists, it is increasingly likely that in the future students will need to graduate with the skills needed to help fill this gap in expertise. As well as this, students from a wide range of academic backgrounds (not just computer science and the mathematical sciences) will be expected to have skills and understanding in this field. It is therefore clear that universities will need to adapt to these changing requirements and develop courses and programs to help facilitate this training.

### Overview of Paper

This paper presents an overview of the work carried out within the Department of Computer Science at University College London to develop a cross-discipline undergraduate module on machine learning. Microsoft's machine learning platform – Azure ML Studio – was used extensively to "bootstrap" the learning process and allow students with little technical background to immediately use and test machine learning models. This constructionist approach is outlined in the next section through three key pillars – scaffolding, discovery, and innovation. The conclusions drawn from this work show the effectiveness of using Azure ML to quickly engage students with complex topics and paves the way for such an approach to be rolled out across a wider population of students.

# CONTEXT

This section outlines the context within which the work was carried out. In particular, a description of the program within which the module is run is given. As well as this, a brief overview of the underlying pedagogical approach is discussed.

## Module Overview and the Integrated Engineering Program

*Machine Learning and Neural Computation* is a final year undergraduate module run by the Department of Computer Science at University College London as part of the Institute of Engineering's Integrated Engineering Program (UCL Engineering, 2017). The course is run over a single academic term from October to December and consists of 20 lectures run over 10 weeks. Although run by the Department of Computer Science, the module is open to enrollment by any student within the Faculty of Engineering. The challenges that this open approach brings are discussed in the following sections.

## A Constructionist Approach to Learning: Scaffolding, Discovery, and Innovation

Problem based learning (Wood, 2003) is a pedagogical methodology whereby students perform independent, self-directed learning through solving an open-ended problem. Within the Department of Computer Science at University College London, such an approach has been refined through a constructionist framework centered upon three steps: scaffolding, discovery, and innovation (Mohamedally, 2012).

The purpose of the scaffolding, discovery, and innovation (SDI) approach is to foster a spirit of invention and enthusiasm within computer science education. Over the past few years, SDI has been primarily applied to the department's innovative second year Systems Engineering module (UCL Computer Science, 2017). This module sees all undergraduate students within the department work for industrial clients to help solve real world business problems (IXN, 2017). These problems are deliberately open ended and allow the students space to discover possible solutions and produce innovative results.

When SDI is applied to module design, the student's learning is split between the lecture hall - where key theories and concepts are taught (i.e. scaffolding) - and the lab classes, homework, and assessment - where students seek to apply this knowledge to open problems (i.e. discovery and innovation). A description of how this approach was followed within the context of the above mentioned machine learning module is given in the next section.

# A BOOTSTRAPPED APPROACH TO TEACHING MACHINE LEARNING

Machine learning is an area of computer science and mathematics that is concerned with enabling computers to learn from experience. Due to the highly technical nature of the field, study of machine learning is typically restricted to advanced level undergraduate or postgraduate level computer science programs. To fully engage with the subject area students need a solid understanding of mathematics along with computer programming experience to build and test the machine learning algorithms being taught. For the module in question this posed a large problem as many of the students enrolled in the course had no programming experience. It was outside of the scope of the module to teach students how to program and the time taken to do so would have meant less time spent on specific machine learning problems.

The solution to the above problem was to use Microsoft's Azure ML Studio platform to "bootstrap" the learning process. Here, bootstrapping refers to the ability to start using the machine learning algorithms being taught with little programming experience. This bootstrapping fits within the general constructionist framework of the module which seeks to foster a spirit of invention and interest through three key processes – scaffolding, discovery, and innovation (Mohamedally, 2012).

**Table 1: Overview of module contents**

| Scaffold Concept | Discovery Lab Class |
|---|---|
| Linear Regression<br>Overfitting and Regularisation | ✓ |
| Logistic Regression<br>Probabilistic Reasoning | ✓ |
| Tree Based Learning<br>Ensemble Learning | ✓ |
| Support Vector Machines<br>Kernel Methods<br>Assessing Model Accuracy | ✓ |
| Unsupervised Learning | ✓ |
| The Curse of Dimensionality | ✗ |
| Deep Learning | ✗ |

This section explores these three processes along with the way in which Microsoft Azure ML Studio was used to bootstrap the learning process.

## Scaffolding: Learning the rules and approaches

The purpose of scaffolding is to teach the base principles and theoretical underpinnings of machine learning. This was done primarily through bi-weekly one hour lectures. Each lecture focused on the theoretical aspects of a particular topic with separate lectures being given to recurring problems within the field (e.g. overfitting, the curse of dimensionality, and assessing model performance). Table 1 details the topics covered in the lectures (referred to as scaffold concepts).

The format of the lectures is to provide a rigorous theoretical understanding of the different algorithms and approaches in machine learning. Little to no attention is given to how these algorithms can be implemented; instead, by focusing on the theory, understanding of how this theory can be applied is left for the students to discover.

## Discovery: Where bootstrapping begins

At the heart of the module is the element of discovery – where students apply the scaffold concepts through guided problem based learning. One of the main obstacles to discovery are the pre-requisites inherent to most technologies used to build and test machine learning solutions. For example, many existing University machine learning modules that contain a practical element make use of programming environments such as MATLAB® or programming languages such as Python. However, due to enrolment being open to any student from the Faculty of Engineering, the course at UCL contains students with little to no programming background. Therefore, if these tools were used then students would need to spend considerable time at the start of the course learning the fundamentals of computer programming as opposed to machine learning.

Microsoft Azure ML Studio (Figure 1) was used as the solution to this problem and to enable students to start the discovery process from the very start of the course. Azure ML Studio provides a simple interactive workspace to build machine learning models. It provides functionality for all the majority of topics covered in the course without the students needing to learn any programming to use them. Therefore, students can start to build machine learning models without the usual requirement of knowing advanced aspects of computer programming. This bootstrapped approach allows students to explore the scaffold concepts and so help them to solidify their understanding.

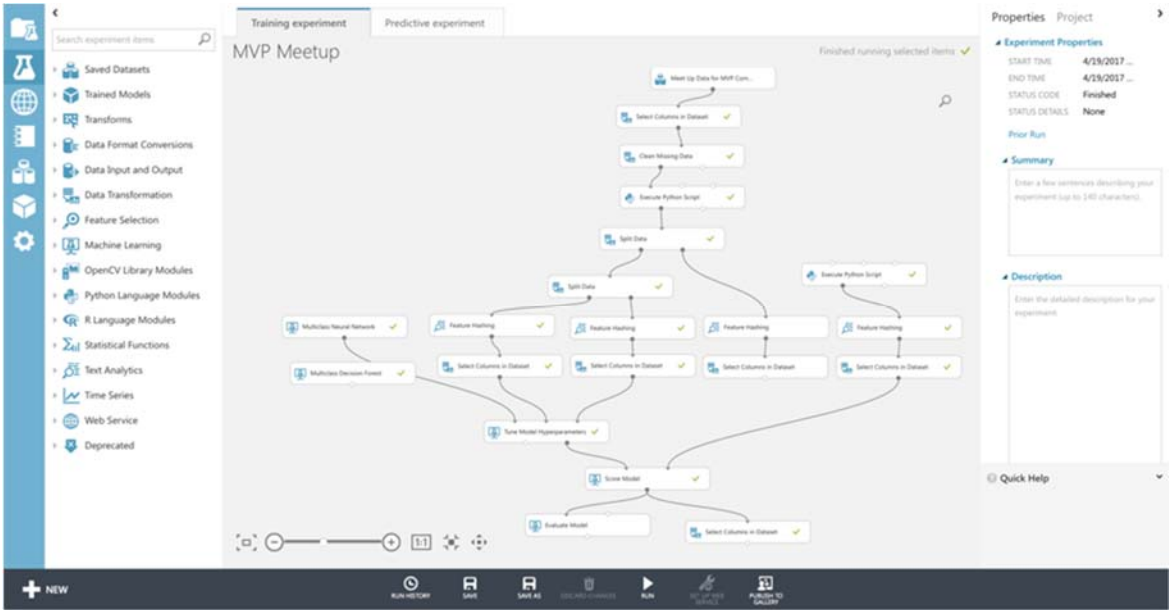


**Figure 1: Microsoft Azure ML Studio provides an intuitive and simple to use interface to quickly build and test machine learning models.**

Discovery mainly takes place through weekly lab classes whose content is aligned to that week's lecture topics (Table 1). The lab classes provide a space for students to explore guided problem based learning. For each lab class, a different data set is given to the students and they are then tasked with using Azure ML Studio to obtain insights and understanding from that data set.

### Innovation: Where scaffolding and discovery leads

Within the SDI framework, innovation is an aspect to be encouraged as opposed to a direct facet of the course (such as the scaffold concepts covered in lectures and the guided discovery of lab classes). However, the course is designed to allow students space to innovate through the assessed coursework.

The coursework for the module is designed in such a way as to encourage the students to seek innovative solutions to a difficult and open ended problem. For the 2016-2017 academic year, the students were tasked with building a machine learning model to predict the results of football matches on a given weekend in the future. Since this is a non-trivial problem without a perfect solution, the students were encouraged to incorporate multiple data sources, try out different algorithmsm and demonstrate their model's performance on previous season's data.

By using a deliberately vague problem, the students are able to use what they have learnt within scaffolding and discovery to produce innovative solutions. This should be contrasted with some more traditional problems where there is a single "right" answer. In such a case, students would be less focused on innovation and more focused on attempting to produce the "correct" result.

## CONCLUSIONS

This paper briefly outlines the teaching strategy for an undergraduate machine learning course. The course utilizes Microsoft Azure ML Studio to enable students to quickly apply theoretical concepts learnt in lectures to real datasets without the need for prior programming experience (i.e. bootstrapping the learning process). This approach is discussed within the scaffolding, discovery, and innovation framework, which is designed to instill a sense of inventiveness and interest in the subject within students.

The proposed module has been rolled out within University College London's Institute of Engineering with great success. The module has proved to be one of the most popular within the Integrated Engineering Program and initial numbers for the upcoming years shows that this continues to be the case.

Future work will be focused on how such a module could be adapted and used outside of traditionally numerate subjects. There is interest in such a module from departments within the Social Science and Humanities faculties, and adapting the module to such settings will undoubtedly see an even greater reliance upon Microsoft Azure ML Studio to help bootstrap the learning process.

## ACKNOWLEDGMENTS

The author would like to thank Lee Stott and Amy Nicholson from Microsoft UK for their help in integrating Azure ML into the module. Thanks also goes to Dr. Simon Julier, Dr. Dean Mohamedally, Dr. Graham Roberts, and Prof. Steve Hailes for their discussions and help in designing the module.

## REFERENCES

Forrest, D., Davenport, T., Guszcza, J., Lucker, J., Raphael, J., Amjad, A., Gold, S., Katyal, V., and Mueller, B., (2016). *Analytics Trends 2016: The next evolution*. Retrieved 24th May 2017 from https://www2.deloitte.com/us/en/pages/deloitte-analytics/articles/analytics-trends.html

IXN (2017). *UCL Industry Exchange Network*. Retrieved 24th May 2017 from http://www.ixn.org.uk

Lohr, S. (2004). Microsoft, Amid Dwindling Interest, Talks Up Computing as a Career. *New York Times*. Retrieved 24th May 2017 from http://www.nytimes.com/2004/03/01/business/microsoft-amid-dwindling-interest-talks-up-computing-as-a-career.html

Marr, B. (2010) Big Data: 20 Mind-Boggling Facts Everyone Must Read. *Forbes*. Retrieved 24th May 2017 from https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/

Mohamedally, D. (2012). *Scaffolding, Discovery and Gamechanging, a strategic approach to Problem Based Learning with Constructionism in Computer Science.* (Professional Certificate in Teaching and Learning in Higher and Professional Education Dissertation). Higher Education Academy, Institute of Education, London UK.

Pesenson, M. Z., Pesenson, I. Z., and McCollum, B. (2010). *The Data Big Bang and the Expanding Digital Universe: High-Dimensional, Complex and Massive Data Sets in an Inflationary Epoch*. arXiv:1003.0879v1 [astro-ph.IM]

Ransbotham, S. S., Kiron, D., and Prentice P. K. (2015). The Talent Dividend: Analytics talent is driving competitive advantage at data-oriented companies. *MIT Sloan Management Review*.

UCL Computer Science (2017). *COMP204P – Systems Engineering Part I.* Retrieved 24th May 2017 from http://www.cs.ucl.ac.uk/teaching_learning/syllabus/undergrad/204p_systems_engineering_project_1/

UCL Engineering (2017). *The Integrated Engineering Program*. Retrieved 24th May 2017 from http://www.engineering.ucl.ac.uk/integrated-engineering/

Wood, D. F. (2003). Problem based learning. *BMJ* 326:328.