Investigation of online and University courses in Machine Learning

Adriaan Louw (53031377)

March 17, 2018

Contents

1 Abstract

An investigation into the state of online and university courses in machine learning

2 Method

I selected 5 online courses and 10 university courses in machine learning. Mostly preferring courses with well know universities. This list is by no means exhaustive. I then tabulated each topic that each course offered.

3 Results

3.1 Online Courses

3.1.1 "Intro to Machine Learning" from Udacity (Stanford University)

This course contains popular topics which include Decision Trees, Regression, Support Vector Machines and clustering. It is also only one of two courses that include Principal Component Analysis (PCA). The course also covers the use of regression.

It does not however include anything on Neural networks or Perceptrons. The course does cover however cover other topics like Baysian networks. (UDACITY, 2018)

3.1.2 "Learning from Data" from Caltech

This course includes Support Vector Machines, Regression, Neural Networks, Kernel methods and Linear modeling is also included. The study of the VC dimension is also included, this is only found in one other course.

This course is the leanest of all the courses with regards to topics taught. There is no mention of Naive Bayes, Baysian Networks, Perceptrons, Clustering or Decision Trees. (California Institute of Technology, 2018)

3.1.3 "Machine Learning" from edX (Georgia Institute of Technology)

With this course we have popular topics which include Decision Trees. Support Vector Machines, Neural Networks, Regression, Kernel Methods, Bayesian Learning and Clustering. It is also the only course to include Markov Decision Processes. VC Dimensions are also included. Also, Instance Based Learning can only be found in this course.

The only popular topic this course is lacking is Perceptrons.(Class Central, 2018)

3.1.4 "Machine Learning" from edX (Columbia University)

This course has most of the popular topics including Decision Trees, Regression, Support Vector Machines Perceptrons, Bayesian Learning, Kernal Methods and Clustering. It is also one of only 2 courses that include PCA and it is the only course that teaches Hidden Markov Models. This course does lack Neural Networks. (edX, 2018a)

3.1.5 "Introduction to Machine Learning" from NPTEL (Indian Institute of Technology Madras)

This course also includes Decision Trees, Support Vector Machines, Regression, Neural Networks, Perceptrons, Bayesian Learning and Clustering. The only popular topic missing is Kernel Methods. (NPTEL, 2014)

3.2 University Courses

The courses selected are: Stanford (2018), University at Buffalo (2018), Massachusetts Institute of Technology (2017), Princeton (2017), University of California, Los Angeles (2017), Michigan State University (2016), University of Notre Dame (2017), Columbia University (2015), University of Melbourne (2018), University of Washington (2012).

The only topic that is covered by all courses is Neural Networks, but only 7 specify that they do Perceptrons. Some mention that they do Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) like University of Notre Dame (2017), but most only specify the term neural networks and do not elaborate. Deep learning is also mentioned with some, but not consistently.

The second most popular topic was Support Vector Machines (SVM), which is covered by 8 of our 10 universities.

A tie for 3rd place are Decision Trees and Logistic Regression. But this come with a caveat, most courses include some sort of regression but not all specify which kind. For instance, this is the case with University of Melbourne (2018). It might be reasonable to assume that at least Linear regression will be covered at most institutions.

Bayesian Learning of some sort was also popular at all universities. Some institutions mentioned only Naive Bayes where others only specified Bayesian Networks. But they all have to do with using the theories of the Reverend Thomas Bayes.

Now coming to less popular topics. Clustering and Kernel Methods were only part of the syllabus of half the courses. Where topics like Principal Component Analysis (PCA), Hidden Markov Models and Natural Language Programming were even less popular.

4 Conclusion

From the online courses under discussion, we can see that not all topics are equally popular. The only 2 topics were present in each course were Regression and Support Vector machines. Regression usually included at least linear regression, but in some courses Logistic Regression was also covered. Two topics were also found in 4 of our courses. These are Decision Trees and Clustering. To round off the popular topics, we have the topic that were included in at least 3 courses. That would be Kernel Methods, Neural Networks and Bayesian Learning.

Some unpopular topics include Hidden Markov Models and PCA. While topics like Q-Learning or Natural Language Processing did not feature at all. From this small dataset it seems like most courses agree on which topics are most important.

As for the university courses, spread of topics covered by our 10 selected courses varied quite a bit. Only Neural Networks, Support Vector Machines and Regression of some kind we consistently covered. The domain space for machine learning seems to be too large to fit everything into a standard university course.

There are a number of skills that will be required in this course.

Firstly a good grounding in statistics is essential. These skills can be learnt online with free courses like Udacity (2018) or edX (2018b).

Secondly, making the reasonable assumption that we will be implementing machine learning algorithms, learning the programming language Python, would be beneficial. python.org (2018) or python guide.org (2018) may be good starting points.

Some elements of Information theory could also be useful. Especially with decision trees and the ID3 algorithm. A nice place to start with this is wikipedia.org (2018).

In general most topics in a course in machine learning should be understandable by a honours student if he or she has studied these additional topics.

References

California Institute of Technology. (2018). *Learning from data.* Retrieved from http://work.caltech.edu/telecourse.html

Class Central. (2018). *edx:machine learning.* Retrieved from https://www.class-central.com/course/edx-machine-learning-8995

Columbia University. (2015). *Machine learning 4771.* Retrieved from http://www.cs.columbia.edu/ jebara/4771/notes/class1x.pdf

edX. (2018a). *Machine learning.* Retrieved from https://www.edx.org/course/machine-learning-columbiax-csmm-102x-3

edX. (2018b). *Statistical modeling and regression analysis.* Retrieved from https://www.edx.org/course/regression-analysis

Massachusetts Institute of Technology. (2017). *6.036.* Retrieved from http://courses.csail.mit.edu/6.036/spring_2017/lectures.html

Michigan State University. (2016). *Cse491.* Retrieved from https://github.com/jiayuzhou/CSE491-2016Fall

NPTEL. (2014). *Introduction to machine learning.* Retrieved from https://onlinecourses.nptel.ac.in/noc18_cs26/preview

Princeton. (2017). *Cos402.* Retrieved from https://www.cs.princeton.edu/courses/archive/fall16/cos402/

python guide.org. (2018). *Learning python.* Retrieved from http://docs.python-guide.org/en/latest/intro/learning/#the-python-tutorial

python.org. (2018). *Python for beginners.* Retrieved from https://www.python.org/about/gettingstarted/

Stanford. (2018). *Cs229.* Retrieved from http://cs229.stanford.edu/syllabus.html

UDACITY. (2018). *Intro to machine learning.* Retrieved from https://www.udacity.com/course/intro-to-machine-learning--ud120

Udacity. (2018). *Intro to statistics.* Retrieved from https://www.udacity.com/course/intro-to-statistics--st101

University at Buffalo. (2018). *Cse474.* Retrieved from https://www.cse.buffalo.edu// chandola/teaching/machinelearningdocs/syllabus.pdf

University of California, Los Angeles. (2017). *Cs188.* Retrieved from http://web.cs.ucla.edu/ sriram/courses/

cs188.winter-2017/html/index.html

University of Melbourne. (2018). *Comp30027.* Retrieved from https://handbook.unimelb.edu.au/subjects/comp30027

University of Notre Dame. (2017). *Cse40625.* Retrieved from https://github.com/cse40625/cse40625/blob/master/syllabus.pdf

University of Washington. (2012). *Css490.* Retrieved from http://courses.washington.edu/css490/2012.Winter/ CSS%20490-590%20-%20Introduction%20to%20Machine%20Learning.html

wikipedia.org. (2018). *Information theory.* Retrieved from https://en.wikipedia.org/wiki/Information_theory