Probabilistic Machine Learning

Prof. Alex Rogers
Department of Computer Science, University of Oxford, UK

Course Description

Probability theory provides a principled foundation that allows computers to reason about uncertainty. It allows us to build machine learning algorithms that perform well, even when we have limited data, by exploiting our understanding of the underlying processes that generated that data. Until recently applying probability theory to such machine learning tasks required knowledge of advanced mathematics and specialist programming skills. However, the advent of powerful probabilistic programming libraries in easy to use languages such as Python has changed this. This course will provide students with a good understanding of basic probability theory and the skills and experience to apply probabilistic programming to solve real-world machine learning and data analysis tasks.

Goals

1. To understand the fundamentals of probability theory, Bayes rule and Bayesian inference.
2. To understand how to use PyMC3, a powerful probabilistic programming framework based on Python, to build sophisticated probabilistic models.
3. To develop experience applying these models to solve real-world machine learning and data analysis tasks.

Prerequisites

Proficiency in high-school level maths and a basic understanding of computer programming (preferably in Python).

Course Materials

There are two recommended textbooks for the course:

I. **Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference**. Cameron Davidson-Pilon, Addison-Wesley Data & Analytics, Oct 2015.

II. **Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ**. Osvaldo Martin, Packt Publishing, Dec 2018.

Either is suitable as a reference. The lecture material will cover all the necessary topics to complete the course.

List of Topics

The course will cover probability theory including common probability distributions, Bayes rule and Bayesian inference. It will introduce Python and the PyMC3 probabilistic programming framework, and will describe how these can be used to solve real-world machine learning and data analysis tasks.

Research Project

All students will be required to work on a research project that will apply the approaches described above to real-world machine learning and data analysis tasks using real data from scientific and engineering examples.

Course Schedule

The first four weeks consist of the following lectures:

I. **Lecture 1: Introduction to Probability Theory**

   The first lecture will introduce the fundamentals of probability theory including conditional and joint probabilities, Bayes rules and the difference between frequentist and Bayesian interpretations of probability.

II. **Lecture 2: Probability Distributions and Bayesian Inference**

   The second lecture will continue the discussion of probability theory to include more advanced topics including as common discrete and continuous probability distributions, Bayesian inference, and conjugate priors.

III. **Lecture 3: Introduction to Python Programming and PyMC3**

   The third lecture will provide a brief overview of programming in Python and will describe the use of probabilistic programming, using PyMC3, to reproduce the results of the previous theory sessions.

IV. **Lecture 4: Solving Real-World Problems with Probabilistic Programming**

   The final lecture will look at how PyMC3 can be used to build sophisticated probabilistic models that address real-world machine learning and data analysis tasks.

The next four weeks will consist of a research project. Each team will be supplied with a real-world dataset from a scientific or engineering application and a description of the underlying process that are believed to have generated the data. Each team will be required to build a probabilistic model that explains the data and to use that model to identify some key parameters relevant to the application.