

Statistical computing with Scala



Level: Advanced

Duration: undefined hours

We are very happy to announce that Prof Darren Wilkinson is running a series of four courses for data science and statistics with Scala.

Course 4 will be concerned with the use of Scala for the development of non-trivial statistical applications. We will see how to exploit non-uniform random number generation and matrix computations in Breeze. Both maximum-likelihood and Bayesian statistical inference algorithms will be considered. Monte Carlo methods for simulation and inference will be examined, in addition to optimisation algorithms. As time permits, we will also discuss more advanced FP concepts, such as type-classes, higher-kinded types, monoids, functors, monads, applicatives and streams, and see how these enable the development of flexible and scalable applications in strongly-typed functional languages.



Course Outline

Course 4 will be concerned with the use of Scala for the development of non-trivial statistical applications. We will see how to exploit non-uniform random number generation and matrix computations in Breeze. Both maximum-likelihood and Bayesian statistical inference algorithms will be considered. Monte Carlo methods for simulation and inference will be examined, in addition to optimisation algorithms. As time permits, we will also discuss more advanced FP concepts, such as type-classes, higher-kinded types, monoids, functors, monads, applicatives and streams, and see how these enable the development of flexible and scalable applications in strongly-typed functional languages.

This suite of 4 half-day courses is aimed at statisticians and data scientists already familiar with a dynamic programming language (such as R, Python or Octave). Scala is a free modern, powerful, strongly-typed, functional programming language. It is fast and efficient, runs on the Java virtual machine (JVM), and is designed to easily exploit modern multi-core and distributed computing architectures. Scala is a favourite language for data engineering teams and others wanting to work with data at scale in an efficient, safe and timely fashion. For similar reasons, it is also very well suited to the development of robust data science, machine learning and statistical applications.

The courses can be taken independently, but do have pre-requisites which are detailed within the Prior Knowledge summaries. They will be delivered through a combination of lectures, live demos and hands-on practical sessions. The courses will be delivered by Prof Darren Wilkinson, a well-known expert in computational Bayesian statistics and a leading proponent of the use of strongly-typed FP languages (such as Scala) for scalable statistical computing. Participants will be expected to use their own laptops and to have a recent version of Java pre-installed. Other set-up instructions will be provided in advance to registered participants.

Prior Knowledge

Course 4 covers relatively advanced material. It assumes a basic familiarity with essential concepts in statistical computing, in addition to a working knowledge of Scala and its use for data science, broadly equivalent to that provided by Courses 1 and 2 ([Introduction to Scala and functional programming](#) & [Scala for data science and machine learning](#)). Note that knowledge of Apache Spark (covered in Course 3) is not necessary for this course.

Contact

hello@jumpingrivers.com