

Swissquote Conference 2018 on Machine Learning

Kay Giesecke (Stanford University)

Title: Towards Explainable AI: Significance Tests for Neural Networks

Abstract: Neural networks underpin many of the best-performing artificial-intelligence systems, including speech recognizers on smartphones or Google's latest automatic translator. The tremendous success of these applications has spurred the interest in applying neural networks in a variety of other fields. In finance, researchers have developed several high-impact applications in risk, investment, and operations management. However, the difficulty of interpreting a neural network model has slowed the implementation of these applications in financial practice, where regulators and other stakeholders insist on model explainability. In this paper, we tackle this issue by developing statistical tests to assess the significance of the input variables of a neural network. We propose a gradient-based test statistic and study its asymptotics using nonparametric techniques. The tests enable one to discern the impact of individual variables on the prediction of a neural network. Experiments using real estate transactions data illustrate their properties. This is joint work with Enguerrand Horel (Stanford).