



Contribution ID: 133

Type: **not specified**

Active learning and adaptive sampling for regression data streams

Wednesday, 18 September 2024 11:30 (30 minutes)

As businesses increasingly rely on machine learning models to make informed decisions, developing accurate and reliable models is critical. Obtaining curated and annotated data is essential for the development of these predictive models. However, in many industrial contexts, data annotation represents a significant bottleneck to the training and deployment of predictive models. Acquiring labelled observations can be laborious, expensive, and occasionally unattainable, making the limited availability of such data a significant barrier to training machine learning models suitable for real-world applications. Additionally, dealing with data streams is even more challenging because decisions need to be made in real-time, compounded by issues like covariate shifts and concept drifts. In this presentation, we will discuss the use of active learning and adaptive sampling techniques to effectively manage label scarcity in supervised learning, particularly in regression data streams. This talk will provide a comprehensive overview of these techniques, followed by detailed discussions on two specific approaches. First, we will dive into stream-based active learning, which aims to minimise labelled data requirements by strategically selecting observations. The focus will be on linear models, and we will explore the impact of outliers and irrelevant features in the sampling process. Next, we will address concept drift monitoring and adaptive sampling, presenting a method to optimise data collection schemes in scenarios where the relationship between input features and the target variable changes over time. The aim of the presentation is to provide an overview of these sampling techniques while highlighting potential applications in real-time data stream scenarios, laying the groundwork for future work in this growing research area.

Type of presentation

Talk

Classification

Both methodology and application

Keywords

Active Learning, Adaptive Sampling, Data Streams

Primary author: CACCIARELLI, Davide (Imperial College London)**Presenter:** CACCIARELLI, Davide (Imperial College London)**Session Classification:** JQT/Technometrics/QE invited session