

Democratizing ML for all scientists: Cloud based machine learning in science

Ben Thurgood and Kevin Jorissen

Scientists across disciplines have turned towards the public cloud for ML/AI based work for several reasons. First, the availability of hardware resources such as GPU servers and clusters, at whatever scale, at any time, allows research questions to be answered more quickly and at larger scale than before. Second, the availability of large data repositories allow new applications both at the training and prediction level. Third, a variety of ML services gives users building blocks to put together robust ML applications with a much lower learning curve, thus democratizing ML. These building blocks range from ready-to-go servers running optimized tensorflow and Mxnet frameworks to black-box speech or image recognition services. The Amazon Sagemaker service occupies the sweet spot for many non-CS scientists: it presents a familiar notebook based environment that allows interacting with data using python code; but is backed by the power of the cloud to automatically run massive training jobs - thereby putting real-world, publishable accuracy within reach of every scientist. It also simplifies and automates the process of creating a labelled data set from scratch saving valuable research time. In this talk we will present the applications and capabilities of these technologies.