Video Tutorial: Overview of AWS SageMaker

“Amazon SageMaker is a fully managed machine learning service. With Amazon SageMaker, data scientists and developers can quickly and easily build and train machine learning models, and then directly deploy them into a production-ready hosted environment. It provides an integrated Jupyter authoring notebook instance for easy access to your data sources for exploration and analysis, so you don't have to manage servers. It also provides common machine learning algorithms that are optimized to run efficiently against extremely large data in a distributed environment. With native support for bring-your-own-algorithms and frameworks, Amazon SageMaker provides flexible distributed training options that adjust to your specific workflows.”

As per Amazon’ belief, “Machine learning in the hands of every developer and data scientist”, Amazon launched a lower level, general level purpose service – “AWS Sagemaker”. It is a platform made for common Machine Learning workflows.

To begin with AWS SageMaker, you spin up a notebook instance which is a Jupiter Notebook application. However, you need not connect with the instance but through AWS SageMaker. One cannot connect with the instance even if they wanted. Jupiter Application supports lots of environment including TensorFlow and Apache MXNet out-of-the-box. It also comes with some built-in algorithm, for instance, PCA, K-Means and XGBoost. You can even use Apache Spark to pre-process the data. One can also use their own framework or other programming language. Later isn’t very simple. It containerize it by packaging them in to Docker application and plug into SageMaker’s training-serving pipeline.

All the data of AWS SageMaker resides in AWS S3 bucket. Hence you need to be sure that S3 bucket lies in the same region as Jupiter instance.

The videos showcases the features of AWS SageMaker by running a small Multiclass classification with Amazon SageMaker XGBoost Algorithm, Apache MXNet Module API and bring your own model – k-means application on AWS SageMaker including Ingesting and preprocessing Data, Training the model and deploying the model.

The biggest advantage of AWS SageMaker is that it supports logging and monitoring without any additional cost.

Total time: 1 hour