Amazon Web Services - An Exploratory Approach

Hang Nguyen & Rosy Lalhlimpuii

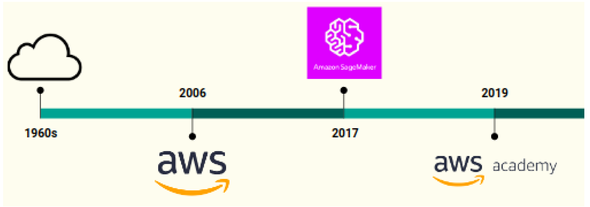
University College Dublin
Ireland's Global University

Abstract

As one of the fastest growing services of the cloud computing providers, Machine Learning (ML) and Artificial Intelligence (AI) platforms are available through diverse offerings: Cognitive Services, ML Platform as a Service - (ML PaaS) and ML Infrastructure Services. Among those, ML PaaS (e.g.Amazon SageMaker, Microsoft Azure ML Services, Google Cloud ML Engine) provides pre-configured and ready-to-use data science environments in which data scientists and engineers could process, train, tune, and deploy the model in a single platform. This thesis explores Amazon SageMaker by comparing among 3 approaches - Studio, Notebook Instance, Console - in 2 use cases - Autopilot (Piali Das 2020), DeepAR algorithm (Salinas et al. 2017).

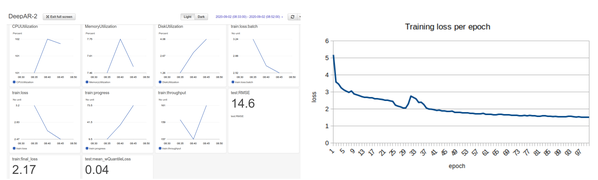
Keywords: AWS, Amazon SageMaker, AutoML, cloud computing, machine learning.

Introduction

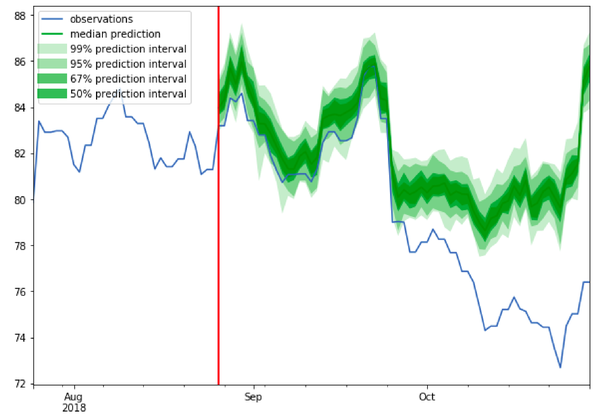


Amazon Web Services (AWS) is the comprehensive and secure cloud platform, offering over 175 fully-featured services with pay-as-you-go pricing from data centres globally.
Amazon SageMaker is the ML platform on AWS that provides infrastructure to build, train, and deploy ML models quickly.
AWS Academy is a program that offers cloud computing curriculum, preparing students to pursue industry-recognized certifications, keep up with AWS Cloud innovation and skills.

Use Case 2: Stock Price Prediction with DeepAR algorithm



		Notebook Instance	SageMaker Console
1	RMSE	12.01	14.6
2	Code	• conda-installed • other data science packages	• process, train and deploy with one click
3	Research workflow	• data exploration • execute and share 'technical' papers	• create Lifecycle Configuration, IAM, VPC, and label data
4	Production workflow	• need advanced infrastructures (e.g. Netflix) to use at scale	• suit for script mode, docker container when using at scale



Limitation/Recommendation

Limitations of SageMaker

- Complex connection with other AWS services.
- Local mode is only available in Notebook Instance, not in Studio Notebook.
- Data pre-processing and Cross-validation are not well-supported.
- Detailed permissions system is needed when starting notebook instance and when onboarding to Studio.
- Python SDK is incomplete and some algorithms are not open-sourced

Recommendations for SageMaker practitioners

- In-depth research on other AWS services (S3, ECR, Step Functions, CloudWatch, Glue, Lambda)
- Keep up-to-date with AWS SageMaker newly-launched features and libraries

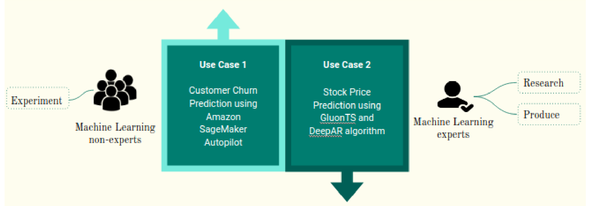
AWS Academy

AWS Academy offers university student with a free and on-demand cloud computing courses that prepares them to pursue certifications and high-demand cloud jobs. As on Aug 2020, the five courses within the AWS Academy platform assists student to achieve the Foundation, Associate level and Machine Learning, Data Analytics Specialty AWS Certifications.

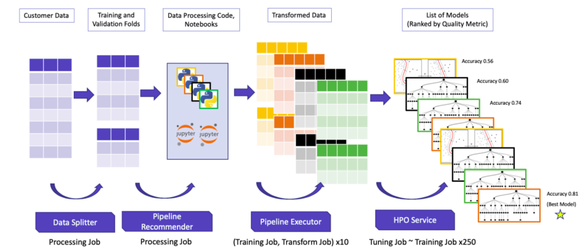
- AWS Academy Cloud Foundations
- AWS Academy Cloud Architecting
- AWS Academy Cloud Developing
- AWS Academy Cloud Operations
- AWS Academy Machine Learning Foundations

Methodology

The thesis explores 3 approaches to Amazon SageMaker in 2 use cases



Use Case 1: Autopilot



		SageMaker Studio	Notebook Instance	Details
1	Accuracy	96.25%	95.65%	
2	Time	✓	✗	Starting a Studio notebook is typically faster than launching an notebook instance
3	Interpretability	✓	✗	- Data Exploration notebook from Studio helps users explore the descriptive statistics with no code needed - Trial Component list allows users to compare the metrics among trial models via CloudWatch
4	Transparency	✓	✗	- Candidate Generation notebook from Studio helps users examine candidates' hyperparameters - Trial Component list: all of model's parameters, configurations and results.

AWS Products - Module Integration

AWS Products	Modules
AWS Cloud Security	Introduction to Business Analytics
AWS Developer Tools	Programming for Analytics
AWS Analytics	Statistics & Simulation Methods
AWS Machine Learning	Data Management & Mining, Statistical Learning
Visual Analytics & Amazon QuickSight	Business Intelligence

Future works

- To cover other AWS approaches (script mode, docker container, AWS Marketplace and local model training).
- To explore Deployment step, including Batch Transform and incremented prediction in real-time, in Use Case 2.

References

Piali Das, N. I. (2020), 'Amazon sagemaker autopilot: a white box automl solution at scale', *International Conference on Management of Data* .

Salinas, D., Flunkert, V. & Gasthaus, J. (2017), 'Deepar: Probabilistic forecasting with autoregressive recurrent networks'.