Intro to Cloud Computing and Machine Learning: Week 8 Notes

Jingchao Zhou
jingchao_zhou@berkeley.edu
Team JSA

Simran Regmi
sregmi@berkeley.edu
Team JSA

1 Introduction to Cloud Computing

Cloud computing is the on-demand delivery of computing services, such as servers, databases, intelligence, over the Internet. It is rapidly growing in popularity, used by countless businesses as well as individuals, and is ultimately making many aspects of the tech field easier. Three of the top companies that offer these services include Amazon(AWS), Google, and Microsoft.

2 Amazon Web Services (AWS)

AWS is an Amazon subsidiary that provides cloud platform services on a pay-as-you-go basis. It is fairly popular amongst all types of consumers due to its scalability and its ability to be easily adapted.

2.1 Current Services Offered

- Amazon EC2: Virtual servers in the cloud
- Amazon Simple Storage Service (S3): Scalable storage in the cloud
- Amazon Aurora: High performance managed relational database
- Amazon DynamoDB: Managed NoSQL database
- Amazon RDS: Managed relational database service for MySQL, PostgreSQL, Oracle, SQL Server, and MariaDB
- AWS Lambda: Run code without thinking about servers
- Amazon VPC: Isolated cloud resources
- Amazon Lightsail: Launch and manage virtual private servers
- Amazon SageMaker: Build, train, and deploy machine learning models at scale

2.2 Pricing Structure

Amazon likes to compare its pricing structure to how one pays for utilities like water and electricity: you only pay for the services you consume, and once you stop using them, there are no additional costs or termination fees. As mentioned earlier, AWS has a pay-as-you-go model which makes it easy for businesses to adapt their services depending on need and not on forecasts.

3 Google Cloud

Google Cloud Platform is a suite of cloud platform services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, file storage, and YouTube. It offers a wide variety of AI/ML resources, allowing businesses to easily implement artificial intelligence and machine learning capabilities.

3.1 Current AI/ML Services Offered

- AI building blocks: Easily infuse AI into applications with custom or pre-trained models.
- AutoML: Custom machine learning model training and development.
- Vision AI: Custom and pre-trained models to detect emotion, text, more.
- Video AI: Video classification and recognition using machine learning.
- Cloud Natural Language: Sentiment analysis and classification of unstructured text.
- Cloud Translation: Language detection, translation, and glossary support.
- Media Translation (beta): Add dynamic audio translation directly to your content and applications.
- Text-to-Speech: Speech synthesis in 220+ voices and 40+ languages.
- Speech-to-Text: Speech recognition and transcription supporting 125 languages.
- Dialogflow: Conversation applications and systems development suite.
- AutoML Tables (beta): Service for training ML models with structured data.
- Cloud Inference API (alpha): Quickly run large-scale correlations over typed time-series datasets.
- Recommendations AI (beta): Deliver highly personalized product recommendations at scale

3.2 Pricing Structure

Google Cloud Platform offers subscription-based pricing, usage-based pricing, and combined pricing options, allowing customers to choose based on the type and scale of the services they need.

4 Microsoft Azure

Microsoft Azure is a cloud computing service created by Microsoft for building, testing, deploying, and managing applications and services through Microsoft-managed data centers. It consists of more than 200 products/services and it popular due its flexibility and security.

4.1 Current AI/ML Services Offered

- Anomaly Detector: Easily add anomaly detection capabilities to your apps
- Azure Cognitive Search: AI-powered cloud search service for mobile and web app development
- Azure Machine Learning: Bring AI to everyone with an end-to-end, scalable, trusted platform with experimentation and model management
- Azure Open Datasets: Cloud platform to host and share curated open datasets to accelerate development of machine learning models
- Computer Vision: Distill actionable information from images
- Content Moderator: Automated image, text, and video moderation
- Custom Vision: Easily customize your own state-of-the-art computer vision models for your unique use case
- Data Science Virtual Machines: Rich pre-configured environment for AI development
- Face: Detect, identify, analyze, organize, and tag faces in photos
- Form Recognizer: The AI-powered document extraction service that understands your forms
- Kinect DK: Build computer vision and speech models using a developer kit with advanced AI sensors

4.2 Pricing Structure

Similar to Google Cloud and AWS, Azure has a pay-as-you-go pricing model which is quite convenient for both businesses and individuals.

4.3 Advantages of Cloud Platforms

- The cloud's pay-per-use model is good for bursty AI or machine learning workloads.
- The cloud makes it easy for enterprises to experiment with machine learning capabilities and scale up as projects go into production and demand increases.
- The cloud makes intelligent capabilities accessible without requiring advanced skills in artificial intelligence or data science.
- AWS, Microsoft Azure, and Google Cloud Platform offer many machine learning options that don't require deep knowledge of AI, machine learning theory, or a team of data scientists.

5 Machine Learning Application

For our learning purposes, we have decided to focus on ML applications using AWS and the cloud platform services it offers. Therefore, we will be focusing on relevant tools and commands as well as introducing ML workflows with AWS.

5.1 Tools and Commands

Amazon Web Services (AWS) offers a large amount of cloud computing services, with applicable use cases ranging from blockchain to to IoT. Some of their more popular products and their uses cases were listed in a prior section; however, we will mainly focus on SageMaker, S3, Rekognition, and EMR.

- **SageMaker:**A fully managed machine learning service that allows data scientists and developers to build and train machine learning models. Some features that are helpful when working with SageMaker are Neo, Ground Truth, and hyperparameter tuning. Neo enables developers to train machine learning models once and run them anywhere in the cloud and at the edge. Ground Truth is a fully managed data labeling service that makes it easy to build highly accurate training datasets for machine learning. Hyperparameter tuning, finds the best version of a model by running many training jobs on your dataset using the algorithm and ranges of hyperparameters that you specify.

- **S3:** A data storage service that has the capability to store the dataset that powers our model training process.
- **Rekognition:** A computer vision software that allows one to identify a person in a photo or video using private repositories of face images or by analyzing a face image against images one has stored for comparison.
- **EMR:** A big data tool that processes vast amounts of data using open source tools such as Apache Spark.

5.2 Introduction to ML Workflows in AWS

There are many AWS cloud computing services that we can utilize to create a machine learning workflow, but we have provided a brief example below to help promote a better understanding of these services.

- take the data from AWS Data Exchange and place -> S3 bucket
- point an AWS Glue Crawler at it to create a Data Catalog of the data
- use Amazon Athena to query, clean, and format the data for training

- load the transformed training set back into S3
- create a Jupyter notebook in Amazon SageMaker to train, deploy, and invoke your predictor