

Machine Learning on the Cloud

Tessema Mengistu (Ph.D.)

mengistu@cs.vt.edu

Outline

- Machine Learning tools and frameworks
- Machine Learning on the Clouds

ML Tools and Frameworks

- Cloud-based ML platforms provide a range of tools and services to support ML workflows:
 - Data preparation
 - Feature engineering
 - Model training, tuning, and deployment

ML Tools and Frameworks

- Tools
 - Jupyter notebook/lab – opensource web application
 - Data cleaning transformation , statistical modelling, visualization, machine learning, etc.
 - Supports over 40 programming languages, including Python, R, Julia, Java, . . .
 - Phyton
 - Pandas – opensource phyton library for data handling and analysis
 - Matplotlib, Numpy
 - Scikit-learn – opensource machine learning library
 - Python programming language
 - Supervised/unsupervised learning

ML Tools and Frameworks

- ML/DL Frameworks

PyTorch

Caffe2

Torch

TensorFlow

Gluon

Chainer

Keras

CNTK

Apache MXNet

ML on GCP

- Google cloud locations:



ML on GCP

- Vertex AI
 - A machine learning (ML) platform
 - Train and deploy ML models and AI applications,
 - Customize large language models (LLMs)
 - Vertex AI provides two options for running your code in notebooks
 - Colab Enterprise:
 - A collaborative, managed notebook environment with the security and compliance capabilities of Google Cloud
 - Vertex AI Workbench:
 - A Jupyter notebook-based development environment for ML workflow.

ML on GCP

- Provides several options for model training and deployment:
 - AutoML
 - Train tabular, image, text, or video data without writing code or preparing data splits
 - Custom Training
 - Gives a complete control over the training process, including using preferred ML framework, writing training code, and choosing hyperparameter tuning options
 - Model Garden
 - Allows to discover, test, customize, and deploy Vertex AI and select open-source (OSS) models and assets
 - Generative AI
 - Gives access to Google's large generative AI models for multiple modalities (text, code, images, speech)
 - Gemini

ML on Azure

- Azure Machine Learning Studio
 - A managed end-to-end machine learning platform for building, fine-tuning, deploying, and operating Azure Machine Learning models, responsibly at scale
 - Supports
 - Notebooks
 - Azure Machine Learning designer
 - Data labeling
 - etc.
- Azure AI Studio
 - Unified platform for developing and deploying generative AI apps and Azure AI APIs responsibly

Amazon SageMaker

- Amazon SageMaker AI
 - A fully managed cloud platform for users to develop ML models from end to end
 - Some of the key features :
 - Data preparation: various tools to preprocess and prepare data
 - Model training algorithms: built-in algorithms for supervised learning, unsupervised learning, and reinforcement learning
 - Model deployment: provides tools for model deployment, either as a batch transform job or a real-time endpoint
 - Scalability:
 - Integration: integrates with other AWS services, such as S3, AWS Glue, and AWS Lambda

Amazon SageMaker



Ground Truth

Set up and manage labeling jobs for highly accurate training datasets by using active learning and human labeling.



Notebook

Provide AWS and SageMaker SDKs and sample notebooks to create training jobs and deploy models.



Training

Train and tune models at any scale. Use high-performance AWS algorithms, or bring your own.



Inference

Create models from training jobs, or import external models for hosting so you can run inferences on new data.



AWS Marketplace

Find, buy, and deploy ready-to-use model packages, algorithms, and data products in AWS Marketplace.

Training on SageMaker

- Amazon SageMaker provides four different ways
 1. SageMaker built-in algorithms
 - Supervised learning
 - eXtrem Gradient Boosting (XGBoost) – opensource implementation of the gradient boost tree algorithm
 - Linear learner
 - Unsupervised
 - K-means
 2. SageMaker supported frameworks
 - Apache MXNet, TensorFlow, PyTorch, SparkML, etc.
 - Offered as per-built Docker images

Training on SageMaker

3. SageMaker custom frameworks
 - User built scripts and algorithms
4. Aws marketplace algorithms
 - Third party ready to use algorithms

Training on SageMaker

Amazon SageMaker built-in algorithms

Classification	Quantitative	Recommendations	Unsupervised	Specialized	
XGBoost	XGBoost	Factorization machines	K-means	Image classification	Neural Topic Model (NTM)
Linear learner	Linear learner		Principal Component Analysis	Sequence-to-sequence	Latent Dirichlet Allocation (LDA)

Training on SageMaker

- Hyperparameter categories:

Model	Optimizer	Data
Help define the model	How the model learns patterns on data	Define attributes of the data itself
Filter size, pooling, stride, padding	Gradient descent, stochastic gradient descent	Useful for small or homogenous datasets

- SageMaker hyperparameter tuning
 - Provides automatic tuning of hyperparameters

Inference on SageMaker

- Deployment
 - Real-time inference:
 - For persistent, real-time endpoints that make one prediction at a time, use SageMaker real-time hosting services.
 - Serverless inference:
 - Workloads that have idle periods between traffic spikes and can tolerate cold starts, use Serverless Inference.
 - Batch transform:
 - To get predictions for an entire dataset, use SageMaker batch transform.
 - . . .

Reference

- <https://docs.aws.amazon.com/whitepapers/latest/aws-overview/machine-learning.html>
- <https://cloud.google.com/vertex-ai/docs/start/introduction-unified-platform>