AZURE ML BOOTCAMP NOTES

June 29th,30th 2023

Notes on Azure ML Bootcamp Session conducted by Mr. Vivek Raj, Azure Tamil Nādu Community
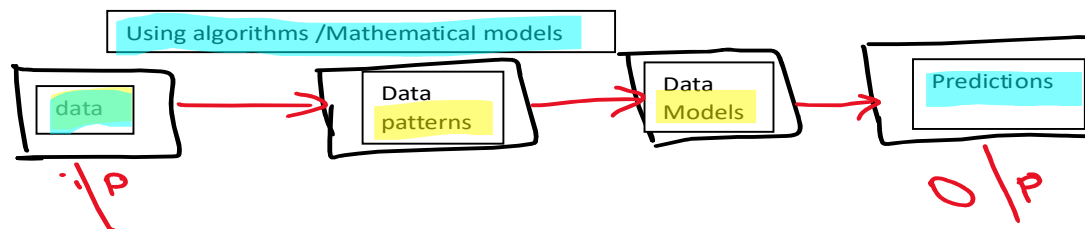
Sridevi
Sridevi.suganya@gmail.com

Table of Contents

Machine learning

- computer learning from data using mathematical models without direct instructions
- a subset of AI
-

Using algorithms /Mathematical models

data → Data patterns → Data Models → Predictions

Types of Machine learning

1.Supervised Learning (input,output)

2.Unsupervised learning (input) – clustering

3.Reinforced Learning (rewards/penalty based learning)

Stages of Machine learning

1.Data Collection & Preprocessing   2.Train the model   3.Validate the model   4.Interpret the results

STAGE 1 : DATA COLLECTION & PREPROCESSING

1. Identify the data source
2. Data collection
3. Data transformation
4. Anomaly detection
5. Clean the data
6. Domain Understanding

STAGE 2 : TRAIN THE MODEL

1. Split the data
2. Select the model
3. Training
4. Hyperparameter - Tuning

STAGE 3 : VALIDATE THE MODEL

1. Validate on the test dataset
2. Evaluate the result
3. Finalize the data model

1. Prediction
2. Monitoring
3. visualizations

Challenges in MLOps

RESOURCES - Creating & mainitaining resources

EXPERIEMENTS - Tracking of ML experiments and its runs

ENVIRONMENTS - Creating /deleting /reusing the environments

MODELS - deploying, creating, mainitaining the different versions

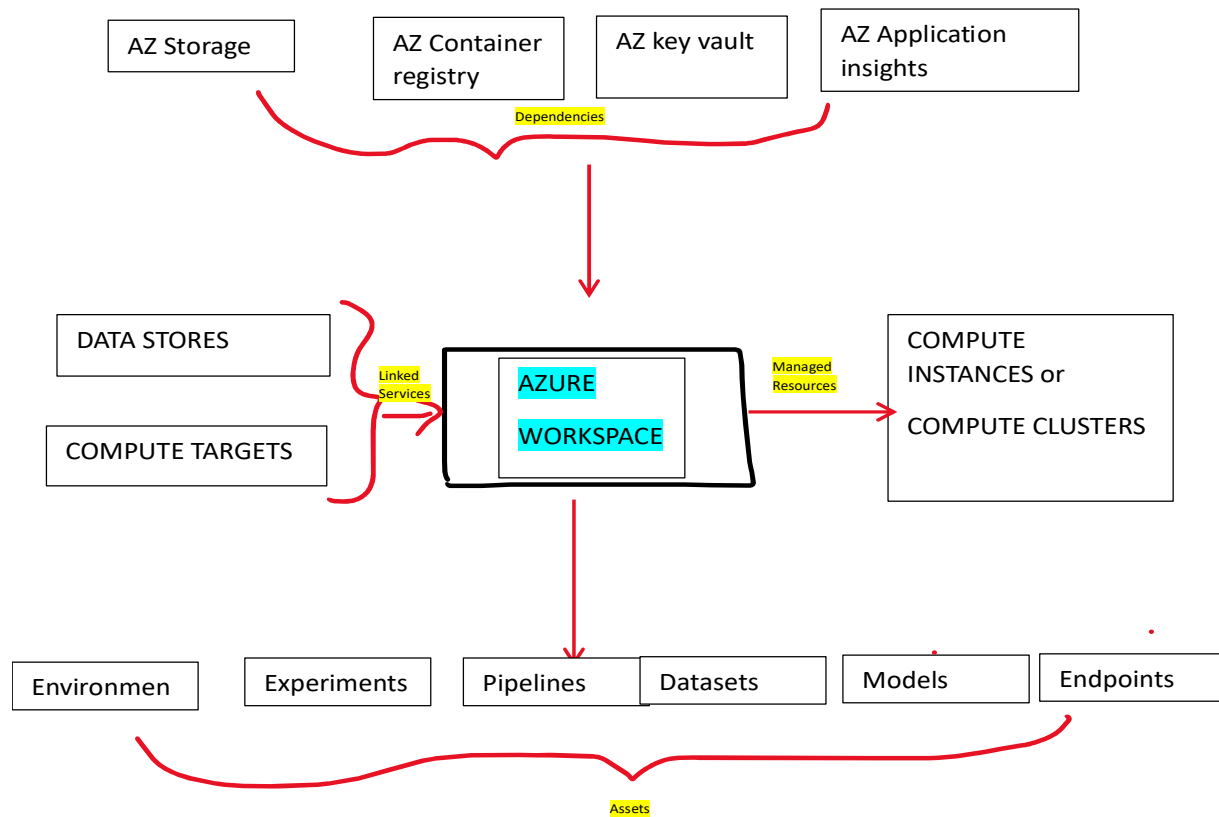**WORKFLOWS** - reusing the workflows

AZURE MACHINE LEARNING

AZURE MACHINE LEARNING

AZURE PORTAL

AZURE CLI

Azure Python SDK

UI +Coding

Complete UI +

Azure ML studio

Azure ML Designer

Azure ML studio

VS code +ML extensions

Other IDE
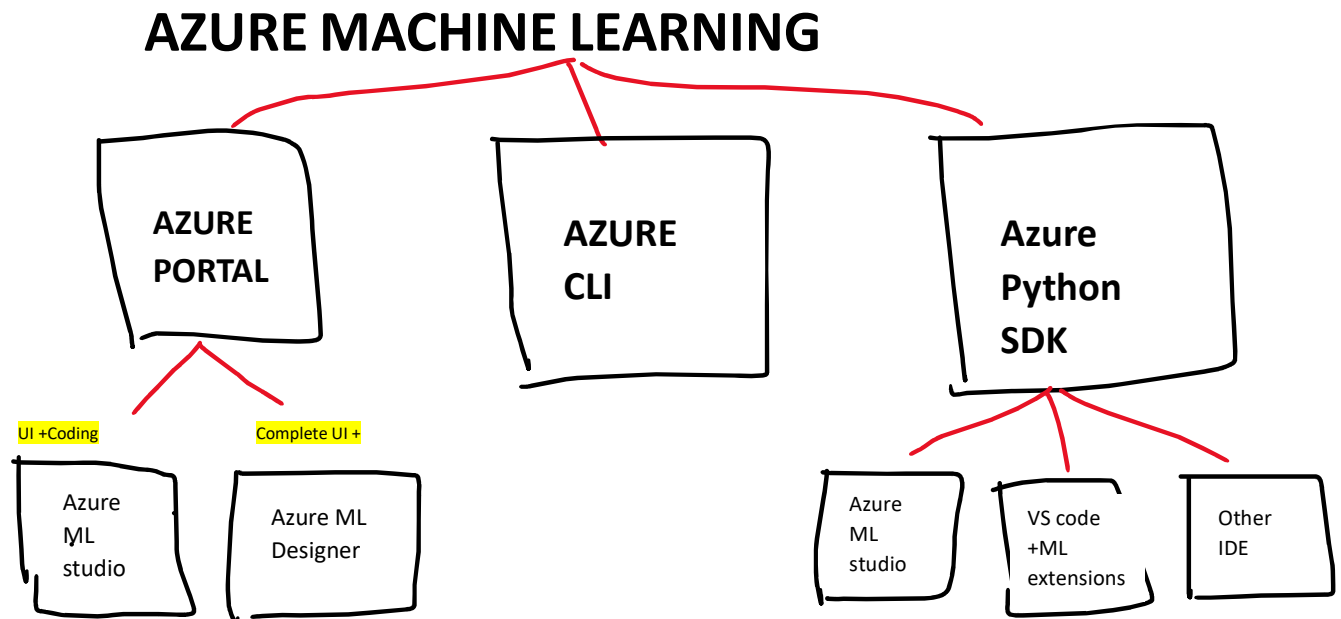
AZURE ML ARCHITECTURE

| AZ Storage | AZ Container registry | AZ key vault | AZ Application insights |

Dependencies

| DATA STORES |
| COMPUTE TARGETS |

Linked Services

AZURE WORKSPACE

Managed Resources

| COMPUTE INSTANCES or COMPUTE CLUSTERS |

| Environmen | Experiments | Pipelines | Datasets | Models | Endpoints |

Assets

Azure ML Workspace

- AZ resource
- organize and coordinates other ML resources(storage , databases, compute)
- assist in execute & share ML workflows.
- ML experimentation, deployment, inference, monitoring of deployed models
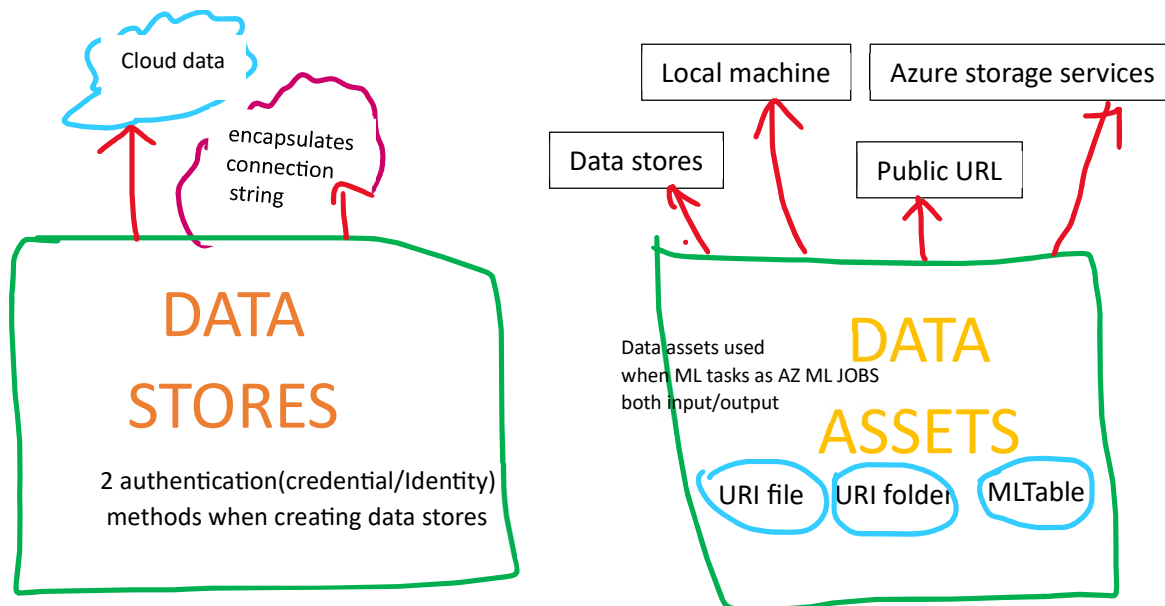
AZURE EXPERIEMENTS

- Named process of running a script or Pipeline
- That can generate output/metrics
- Can be tracked in AZ ML workspace

EXPERIMENT RUN CONTEXT

- after submitting the experiment
- use the experiment run context to INITIALISE & END the experiment
- log, monitor every experiement

STAGE 1 - COLLECTING + PREPROCESSING DATA

Bringing data to Azure

Cloud data

encapsulates connection string

Local machine

Azure storage services

Data stores

Public URL

DATA STORES

2 authentication(credential/Identity) methods when creating data stores

Data assets used when ML tasks as AZ ML JOBS both input/output

DATA ASSETS

URI file

URI folder

MLTable

Authentication Methods in Data stores

1. Credential based Authentication
   a. Service Principal
   b. SAS token
   c. Account key
2. Identity based Authentication
   a. Azure AD identity
   b. Managed Identity

STAGE 2 - TRAIN THE MODEL

Training ML Models in Azure using scripts

From Notebooks to Scripts

Notebooks	Scripts
Ideal for exploring data	Ideal for production loads
Ideal for Prototyping	Integrates with production pipelines
Code development – visual + interactive	consistency
But for production, Notebooks lacks scalablity , automation	reproducibility

Production scripts

- Remove non essential code
- Refactorise as functions
- Test it in terminal

Azure Machine Learning command Jobs

ML command jobs are for running Azure ML production scripts

In command jobs ,

- Configure parameters such as input , output destinations ,compute environment
- Automate jobs
- Schedule jobs
- Control over prod. Environment

5 steps for ML models

1. Create a script
2. Setup ML workspace
3. Define a compute target
4. Configure the command job
5. Submit the command job for experiment

Rate limit errors Vs Capacity errors in GPT

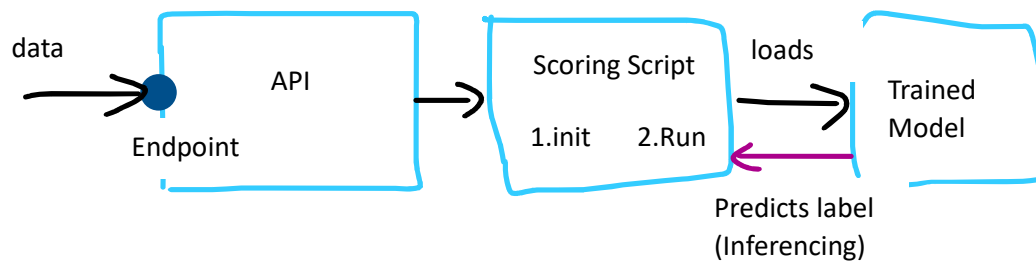Rate Limiting – no: of requests in a particular time

Capacity error – server side request handling

Productionizing the ML Model

Model Management

Endpoints

Endpoint is a specific location within an API(set of rules for communication between devices) that accepts requests and sends back responses

Inferencing

Scoring script

The scoring script is a Python file (. py) that contains the logic about how to run the model and read the input data submitted by the batch deployment executor.

It has init – run

Types of Online endpoints

1. Managed Online endpoints - managed by 3rd party
2. Kubernetes Online endpoints – user manages

Steps for deploying

1. Model assets like Pickle file or Registered model in Azure workspace
2. Scoring script
3. Environment or separate inferencing environment
4. Compute configuration (with Scaling +size)

model pickle file

Pickle is a useful Python tool that allows you to save your ML models, to minimise lengthy re-training and allow you to share, commit, and re-load pre-trained machine learning models.

 Most data scientists working in ML will use Pickle or Joblib to save their ML model for future use.

Automated ML takes care of Feature engineering

Managed Endpoint – no scoring script required

Pipeline Vs Single standalone Job

- Pipeline – preprocessing , training as separate step
- Single standalone job – everything in one go

Executing as Single Standalone Job using script

Use case : Analyse if the client will pay the amount or not

https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset

1. Create AZ machine learning service from azure Portal
2. Download the config.json file from the AZ ML service
3. Login to the Learning Studio and create a ML workspace, Compute resource
4. From compute resource – click Jupyter
5. download the file from github https://github.com/Vivek0712/azuremlbootcamp
6. the config.json file and Github file should be in the same folder.
7. run the file step by step as below :
8. Authentication
9. Connecting Workspace
10. Connecting Data to Azure ML Workspace
    a. Register the dataset to the ML workspace
11. Creating Environment
    a. create a dependencies folder,
    b. write .conda.yaml
    c. register the environment to the ML workspace
12. Create Training
    a. create a src folder
    b. create main.py
    c. Using MLFlow to train the dataset & register the model & save the model
13. Configure & Run command job
14. Deploy to an online endpoint
    a. create a unique name for endpoint
    b. create an online endpoint
    c. pick the latest version of the model

15. Test the deployment
    a. create a sample json file
    b. test using the endpoint,sample json file

Visual representation of the pipelines

Step by step execution

AUTOMATED MACHINE LEARNING

Everything is made automatically - Feature selection / model selection by multiple pipelines

This is for low code data scientists