

Assignment on Amazon SageMaker

Objective:

To understand the end-to-end process of building, training, deploying, and monitoring machine learning models using Amazon SageMaker and its integration with AWS services.

Part A – Short Answer Questions

1. What is Amazon SageMaker and how does it simplify the machine learning lifecycle?
2. Explain the role of Amazon S3 in SageMaker workflows.
3. Differentiate between SageMaker Studio and SageMaker Notebooks.
4. What are the advantages of using built-in algorithms in SageMaker compared to custom algorithms?
5. Define hyperparameters. Why is hyperparameter tuning important?
6. List and explain any three preprocessing steps commonly performed before training a model.
7. What are the different deployment options in SageMaker?
8. Mention at least two AWS services that integrate with SageMaker for security and automation.
9. Explain the importance of model monitoring after deployment.
10. What is Boto3 and how is it used in SageMaker?

Part B – Long Answer Questions

1. Draw and explain the high-level architecture diagram of Amazon SageMaker.
2. Describe the workflow of an ML project in SageMaker from data collection to monitoring.
3. Explain with examples how data preprocessing is performed in SageMaker notebooks.
4. Compare real-time endpoints, batch transform, and asynchronous inference in model deployment.
5. Discuss advantages and challenges of using SageMaker in enterprise-scale ML projects.

Part C –lab 3 and 4

1. Data Ingestion – Upload a dataset of your choice to Amazon S3 and verify the bucket structure.
2. Preprocessing in Notebooks – Load the dataset from S3 and perform preprocessing (missing values, encoding, normalization, splitting).
3. Model Training – Use the XGBoost built-in algorithm to train a model and save outputs in S3.
4. Hyperparameter Tuning – Perform automatic tuning and compare accuracy before and after tuning.
5. Model Evaluation – Evaluate model performance using accuracy, precision, recall, F1-score, or RMSE.
6. Deployment – Deploy the model as a real-time endpoint and test predictions using Boto3.
7. Monitoring – Enable model monitoring and simulate data drift to observe results.

Part D – Case Study

Assume you are working for a financial company that wants to predict whether a customer will default on a loan.

Dataset: Bank loan dataset (features like income, age, credit score, loan amount, etc.).

Tasks:

1. Collect and preprocess data using SageMaker.
2. Train a classification model (e.g., XGBoost).
3. Tune hyperparameters to improve accuracy.
4. Deploy the model and test predictions.
5. Monitor the model for performance drift over time.

Prepare a report with screenshots of each step.

4. How does SageMaker compare with Google Vertex AI or Azure ML Studio?
5. Explain how serverless architecture in AWS Lambda can integrate with SageMaker.