

Machine Learning Modeling Day 3 (Second half) Notes

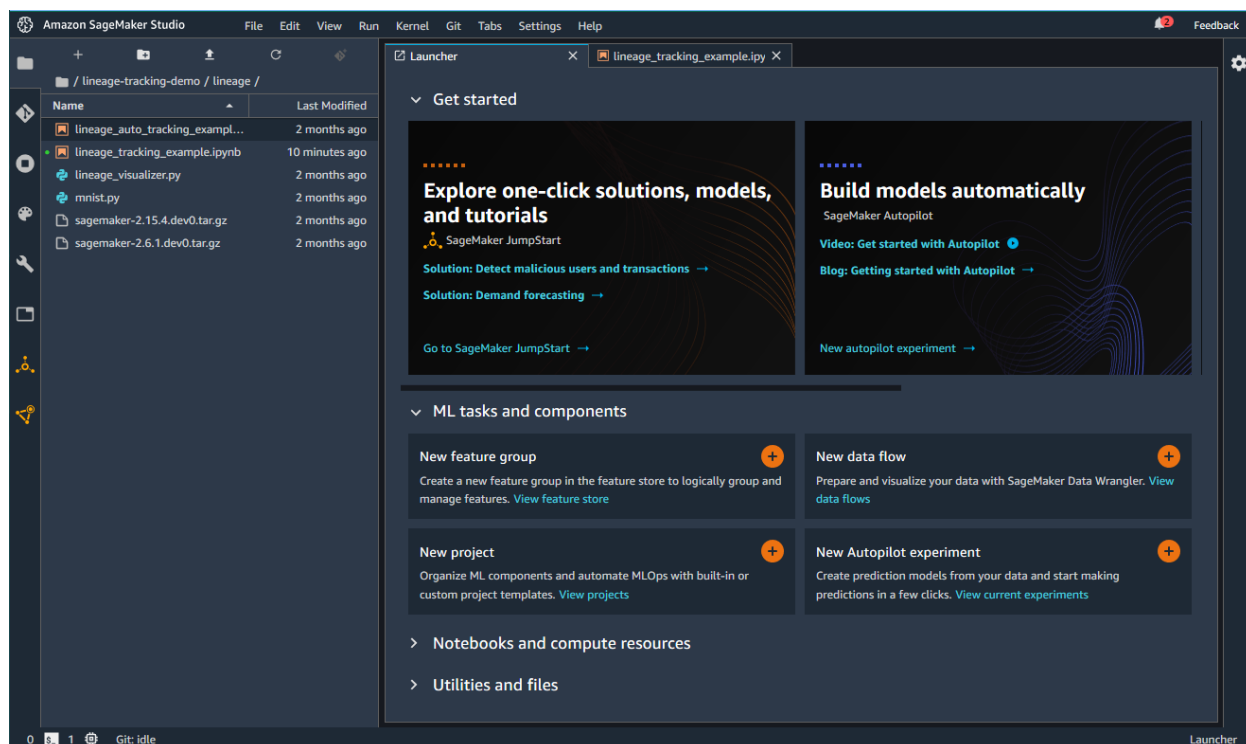
SageMaker Studio

Amazon SageMaker Studio is essentially a interface-based IDE for machine learning (ML) that simplifies many of the necessary and time consuming tasks in the field. Relatively complicated steps such as model deployment can be achieved by simply right clicking and selecting the 'deploy model' button in SageMaker Studio. SageMaker Studio may be seen as a feature-enriched and more efficient Jupyter Lab.










Things to note:

- In SageMaker Studio, kernels are decoupled from the notebook instances, providing more granular control over ML instances that host notebooks.
- Local mode is not supported in SageMaker Studio.
- Upon launching a notebook for the first time, it will take around 3-5 minutes before the kernel starts since the instances along with other resources are still being provisioned.
- Kernels may also be changed like in Jupyter Notebook and Jupyter Lab.
- To support external kernels and custom images like the R kernel, Ruby on Rails, or JavaScript, the user must create their own docker container, push it to Amazon Elastic Container Registry (ECR), create an image via the SageMaker console, and attach it to the SageMaker domain.
- Git settings are also available in the left sidebar.

SageMaker Studio UI



Sidebar components

Icon	Description
	File Browser Choose the Upload Files icon () to add files to Studio. Double-click a file to open the file in a new tab. To have adjacent files open, choose a tab that contains a notebook, Python, or text file, then choose New View for File . Choose the plus (+) sign on the menu at the top of the file browser to open the Studio Launcher.
	Running Terminals and Kernels For more information, see Shut Down Resources .
	Git You can connect to a Git repository and then access a full range of Git tools and operations. For more information, see Clone a Git Repository in SageMaker Studio .
	Commands (Ctrl + Shift + C) The majority of the menu commands are available here.
	Notebook Tools You can access a notebook's metadata through the Advanced Tools section. This icon is displayed only when a notebook is open.
	Open Tabs Provides a list of open tabs, which is useful if you have multiple open tabs.
	SageMaker Jumpstart Provides a list of solutions, model endpoints, or training jobs created with SageMaker Jumpstart.
	SageMaker Components and registries Provides a list of projects, data wrangler flows, pipelines, experiments, trials, models, or endpoints, or access to the feature store.

Key features

SageMaker Experiments

- An tool that aids in locating, organizing, and identifying all the moving parts in an ML experiment.
 - i.e. model lineage, model artifacts, model hyperparameters, etc.

Automated Machine Learning via Autopilot

- Simply provide the S3 bucket address where the input data is stored, optionally specify the target column, and fill in the remaining requirements, and the entire training process will be performed automatically.
- **Candidate Definition Notebook**
 - provides models and algorithms that are viable candidates for the training job



Dataset from S3 should at least have 500 rows in order for Autopilot to be initiated.

- Autopilot will be in charge of performing the ff:
 - Pre-processing
 - Generating Candidate Definitions
 - Feature Engineering
 - Model Tuning
 - Generating *Explainability* Report

Cloud AutoML

- The concept behind automated ML.
- Autopilot is the AWS product of AutoML

ML Libraries and Frameworks with SageMaker

Main workflow:

- Preprocessing
 - SageMaker training
 - SageMaker deployment
 - Inference
1. Transfer external code and/or scripts to a Python script. The end result should be a custom script file that will perform the training job by simply being executed.
 2. Test if it works locally
 3. Instantiate the SageMaker Estimator object and use the train script during the training and deployment stages.
 4. Verify if the inference endpoint is operational.



Everything should work **outside** SageMaker