

AWS のサービスの調査

1. 調査対象サービス

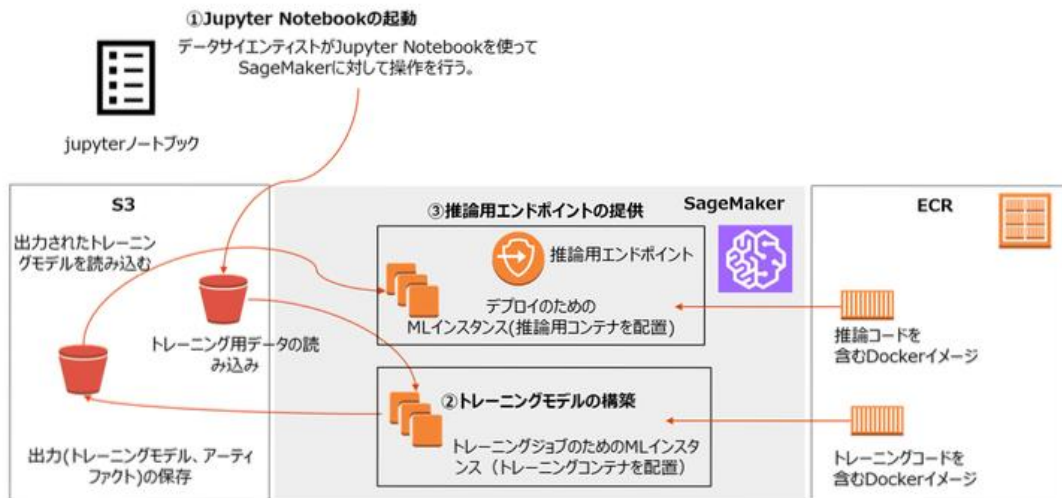
- ✓ SageMaker
- ✓ CodePipeline

1 SageMaker

1.1 サービスの概要

サポートされるワークフロー

- データ収集・準備
- モデル構築
- トレーニング環境の構築・自動化
- モデルのトレーニング及びデバッグ、チューニング
- 実験管理
- 本番環境へのデプロイ
- モデルの監視
- 推論結果の検証
- 本番環境でのスケーリングと管理



SageMakerのアーキテクチャー

特徴

☆ インターフェース

SageMaker では、Jupyter Notebook を使用して、「インスタンスの作成」、「モデル構築」、「トレーニング」、「デプロイ」までのフローを実行することができる。対話型のブラウザで実行することで、実行結果を分かりやすい形で監視することができる。

☆ 主要な機械学習フレームワーク

Tesnsorflow や sklearn などの機械学習モデルを Docker コンテナ上で実行するように構成されている。また、一般的な機械学習アルゴリズムが事前にインストールされており、他の機械学習サービスと比較して、高いパフォーマンスでアルゴリズムを実行できる可能性が高い。

☆ 操作性の高さ

SageMaker のコンソールからワンクリックで高度なトレーニング他チューニングを実行できるようになっている。

☆ フルマネージメントスケーリング機能

インフラ部分については自動的に管理されてスケーリング(規模の増減、面積の拡大縮小) されるので、ペタバイト規模のモデルトレーニングが簡単にスケーリングできる。

1.2 基本的な使い方

- ✧ 専用の S3 バケットを用意
- ✧ ノートブックの作成とインスタンスの指定
- ✧ 実行環境自体はコンテナイメージとして ECR 上に用意
- ✧ 使用したい環境をコンテナイメージから選択して利用する。

1.3 各種サービス



SageMaker Studio

モデルの構築、トレーニング、デプロイに必要な各ステップを管理、可視化できる。

- データのアップロード
- 新規ノートブックの作成
- モデルのトレーニングと調整
- 結果の比較
- 本番環境にモデルをデプロイする

をすべて 1 箇所から実行できる。Jupyter Notebook 上で実行でき、計算リソースのコントロールが可能。ノートブック自体を他者と共有することができる。モデルのパラメータ等の組み合わせを変えて実験する際にはリーダーボードを起動し、実験を自動的に可視化し、追跡、並べ替えやランク付けができる。

SageMaker Studio のノートブックから、Amazon EMR で動作する Apache Spark データ処理環境を、数クリックでブラウザ、発見、接続することができる。

Ground Truth



フルマネージド型のデータラベル付けサービス：

機械学習のための高精度なトレーニングデータセットを簡単に構築することができる。

Amazon SageMaker Data Wrangler

前処理専用のサービス

Amazon SageMaker Feature Store

機械学習のパラメータ等の保存、更新、取得、共有を管理するためのサービス

Amazon SageMaker Debugger

トレーニングメトリックとシステムリソースをリアルタイムで監視して ML モデルを最適化するサービス。学習中の loss などのトレーニングメトリックをリアルタイムで監視、異常を検出した際にはアラートを送信し、モデル予測を即座に修正できる。必要な精度を達成するとトレーニングプロセスを自動的に停止するため、機械学習モデルのトレーニングにかかる時間とコストを節約できる。

Amazon SageMaker Pipeline

- ✧ データの読み込み
- ✧ データの変換
- ✧ トレーニングと調整
- ✧ デプロイ

1 週間に数十個の ML モデルを構築し、大量のデータ、数千回のトレーニング実験、および数百の異なるモデルバージョンを管理できます。ワークフローを共有および再利用してモデルの再作成や最適化を行い、組織全体で ML を拡張できる。

Amazon SageMaker Clarify

さまざまなグループ間でのトレーニングデータまたはモデルの予測動作における不均衡、バイアスを検出し、データとモデルで測定するサービス。モデル入力の重要度を調べて、モデルが予測を行う理由を説明することもできる。指定した属性を調べることでデータの準備中、モデルトレーニング後、デプロイされたモデルで潜在的なバイアスを検出できる。

1.4 その他の機能

Git との統合

SageMaker 上で Git レポジトリを登録しておくことで、ノートブックインスタンス起動時に clone される。

参考文献

<https://www.acrovision.jp/service/aws/?p=1237>

<https://corp.logly.co.jp/blog/417>

<https://www.inoue-kobo.com/aws/sagemaker-with-mycontainer/index.html>

<https://pages.awscloud.com/rs/112-TZM->

[766/images/20201015_AWS%E3%81%A6%E3%82%99%E6%A7%8B%E7%AF%89%E3%81%99%E3%82%8BMLOps%E5%9F%BA%E7%9B%A4_%E9%85%8D%E5%B8%83%E7%89%88_AIMLTokyo7.pdf](https://pages.awscloud.com/rs/112-TZM-766/images/20201015_AWS%E3%81%A6%E3%82%99%E6%A7%8B%E7%AF%89%E3%81%99%E3%82%8BMLOps%E5%9F%BA%E7%9B%A4_%E9%85%8D%E5%B8%83%E7%89%88_AIMLTokyo7.pdf)