AWS Summit

Tokyo

amazon web services

Dev-03:Service

今日から始められる、機械学習！ Amazon Machine Learningのご紹介

Toshiaki Enami, Yuta Imai
Solutions Architect, Amazon Data Services Japan, K.K.

amazon web services

■Gold Sponsors

intel
Empowered by Innovation
NEC
NTT DATA
NetApp
Serverworks
HITACHI
Inspire the Next
TREND MICRO
WORKS APPLICATIONS

■Global Sponsors

infor

■Silver Sponsors

CTC
Challenging Tomorrow's Changes
FPT
JBCC
SKYARCH
あなたの側で、あなた以上に考える。
iSiD
IT Solution Innovator
ナレコムクラウド
NRI 未来創発
Dream up the future.
野村総合研究所
FUJISOFT
amazon web services

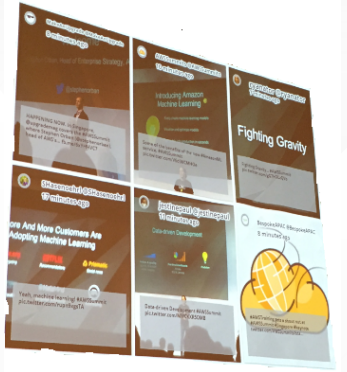
■Bronze Sponsors

■Global Tech Sponsors

■Logo Sponsors

amazon web services

ハッシュタグ **#AWSSummit**
と **#DevCon**で、**皆さんのツイート**
が展示エリアの**大画面に表示されます**

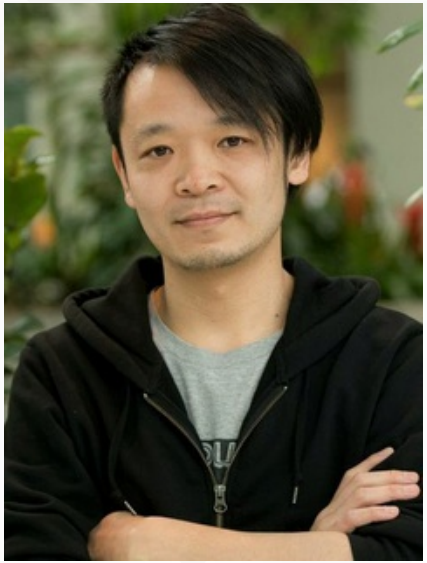


公式アカウント**@awscloud_jp**
を**フォローすると**、ロゴ入り
コースターを**プレゼント**

【コースター配布場所】
メイン展示会場、メイン会場1F受付、デベロッパーカンファレンス会場

amazon web services

自己紹介



今井雄太（いまいゆうた）
yuimai@amazon.co.jp

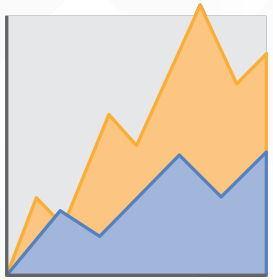
- ソリューションアーキテクト
- ゲーム、広告系のお客様を主に担当
- 好きなAWSのサービス：
  - Amazon Elastic MapReduce

アジェンダ

1. 機械学習とは？
2. Amazon Machine Learning
3. アーキテクチャへの組み込み

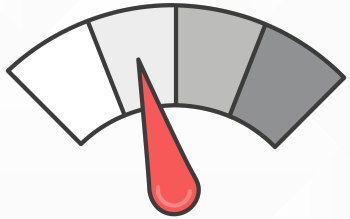
Part1: 機械学習とは？

3種類のデータ駆動型アプリケーション



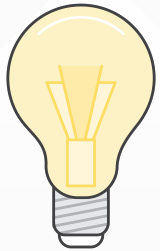
遡及的分析とレポーティング

Amazon Redshift,
Amazon RDS
Amazon S3
Amazon EMR



即時の判断
リアルタイム処理とダッシュボード

Amazon Kinesis
Amazon EC2
AWS Lambda



予測
スマートアプリケーションを作成可能にする

Amazon ML

機械学習の例
- すべてAmazon Machine Learningで実現可能な例-

- このメールはスパムメールか？
  - 過去のメールアーカイブをもとにYes/Noを予測する
- この商品は本、日用品、食品のいずれなのか？
  - 多くの商品データをもとにその商品のカテゴリを予測する
- 明日の売上はどのくらいになるか？
  - 過去の売上データなどをもとに明日の売上を予測する

機械学習の例

- このメールはスパムメールか？
  - 過去のメールアーカイブをもとにYes/Noを予測する

教師データ

送信者domain	サーバーIP	送信時刻	言語	Spam?
A	123.123.123.123	...	JP	1
B	111.111.111.111	...	EN	
B	111.111.111.111	...	EN	
D	123.456.789.012	...	FR	1

別の方法でSpam判定済みの教師データ（過去のデータ）をもとに

予測対象データ

送信者domain	サーバーIP	送信時刻	言語	Spam?
B	123.123.123.123	...	JP	
A	123.456.789.012	...	FR	

予測対象データ（新しいデータ）のSpam判定をする

機械学習の例

- この商品は本、日用品、食品のいずれなのか？
  - 多くの商品データをもとにその商品のカテゴリを予測する

教師データ

商品名	価格	大きさ	メーカー	カテゴリ
A	123	...	AA	Book
B	456	...	BB	Book
C	100	...	CC	Food
D	500	...	DD	Grosery

大量の商品のデータをもとに

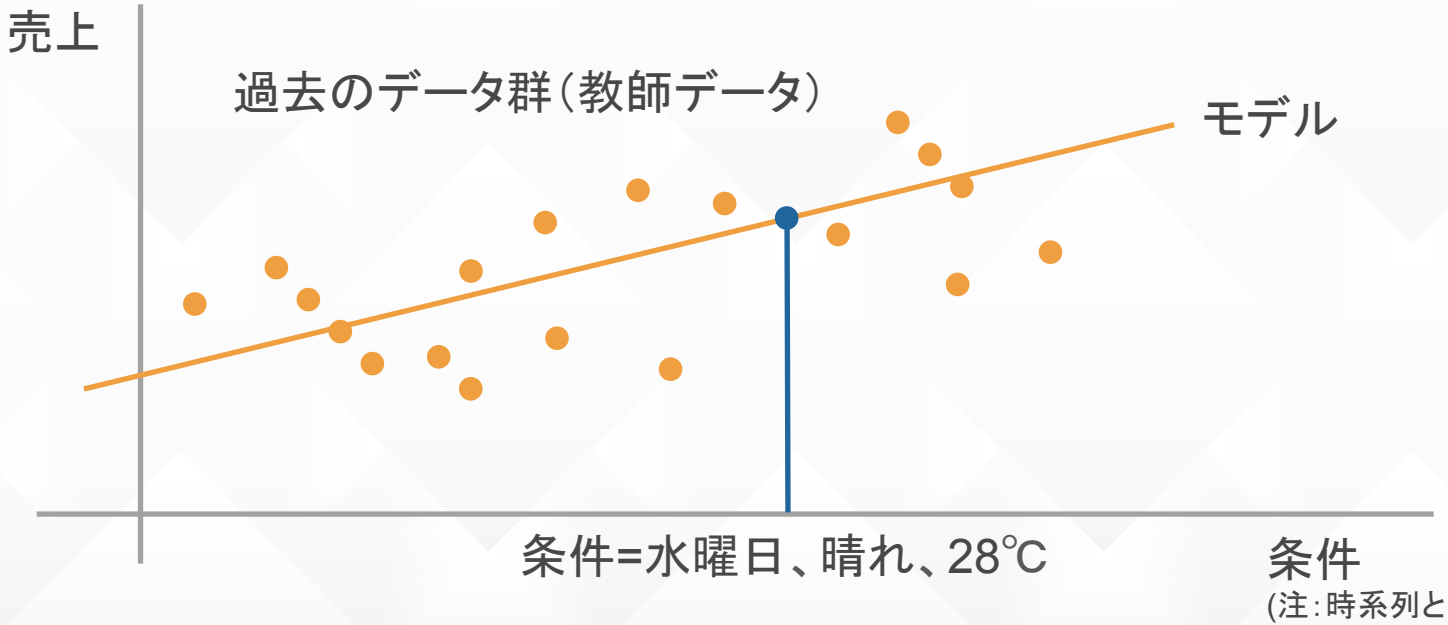
予測対象データ

商品名	価格	大きさ	メーカー	カテゴリ
E	500	...	EE	
F	600	...	FF	

未知の商品のカテゴリを予測する

機械学習の例

- 明日の売上はどのくらいになるか？
  - 過去の売上データなどをもとに明日の売上を予測する



その他の例 …

詐欺の検知	不正取引の検知、不正クリック検知、スパムeメールのフィルタリング、疑わしいレビューのマーキング …
パーソナライゼーション	コンテンツのレコメンデーション、予測的なコンテンツロード、ユーザエクスペリエンスの改善 …
ターゲットマーケティング	オファーとお客様のマッチング、マーケティングキャンペーンの選択、クロスセリングやアップセリング …
コンテンツ分類	ドキュメントのカテゴリー分類、履歴書と採用マネージャのマッチング …
変動予測	サービスを使うのを止めそうなお客様の検知、無料ユーザからアップグレードのオファー …
カスタマーサポート	お客様からのメールの適切な転送先推測、ソーシャルメディアリスニング …

スマートアプリケーション(機械学習を有効活用したアプリケーション)が続々と登場しないのはなぜか?

1. 機械学習の専門家が**少ない**

2. 機械学習の仕組みを作り、スケールさせることは**技術的に困難**

3. モデルとアプリケーションのギャップを縮めるには、**長い時間と高い費用**が必要になる

スマートアプリケーションを作るには

専門家	技術	使いやすさ
データサイエンティストの人数は限られる	多くの選択肢があるが決定的なものが無い	複雑で間違いを起こしやすいワークフロー
外注するのは高くつく	使いこなし、スケールさせることが困難	特殊なプラットフォームとAPI
	カスタムソリューションを作成するために多くの細かい作業が*毎回必要*になる	モデルライフサイクル管理を無駄に再発明

スマートアプリケーションを作るには

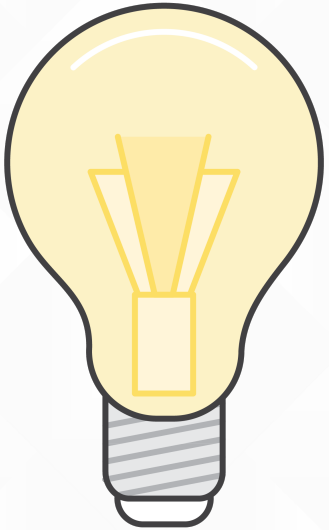
- 機械学習に強くて
- RやPython,場合によってはHadoopやSparkに明るくて
- 特定のビジネス分野の経験が深い

こういう人を採用するのは難しい！

そこで・・・

amazon web services

Amazon Machine Learningの登場

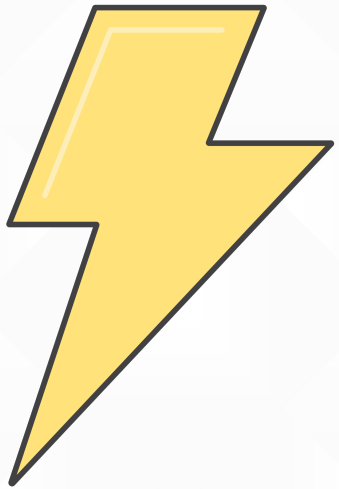


使いやすく、マネージド。開発者のために作られた機械学習サービス

Part2: Amazon Machine Learning

amazon web services

Machine Learning as a Service



Amazonが提供するアルゴリズム

- 利用者は自分でアルゴリズムの実装や詳細なチューニングを行う必要がない

パッケージサービスとしての提供

- 必要なワークフローが予め提供されている

スケーラビリティ

- 利用者はシステムの拡張やその運用についても考える必要がない

Amazon Machine Learningでできること

amazon web services

取り扱える予測モデルとアルゴリズム

二項分類

ロジスティック回帰

このメールはスパム？Yes？No？

多クラス分類

多項式ロジスティック回帰

これは本？車？食べ物？

回帰分析

線形回帰

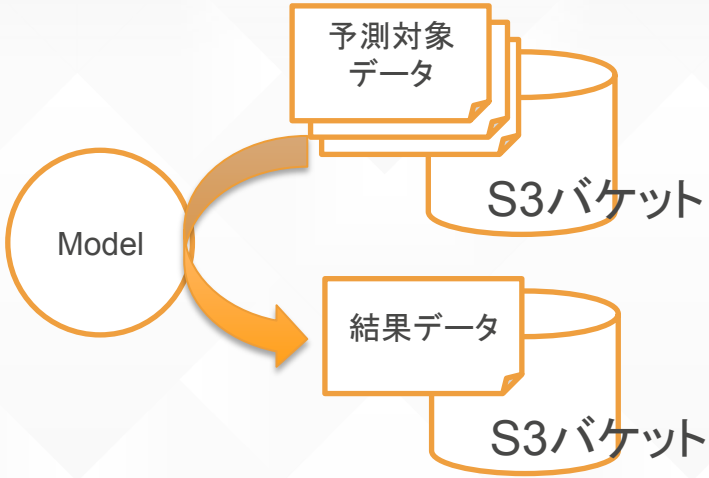
あしたは水曜日。在庫はいくつくらい必要？

これらのモデルを使って・・・

予測手法

バッチ予測

S3等にアップロードされた予測対象データに対してまとめて予測を実施



リアルタイム予測

データ１件ずつAPIを使って予測を実施する

```
ml     = Aws::MachineLearning::Client.new
record = { attr_A: ‘foo’, attr_B: ‘bar’,... }
result =  ml.predict(
    ml_model_id: MODELID,
    record: record,
    predict_endpoint: ENDPOIND
)
```

Amazon Machine Learningの使い方

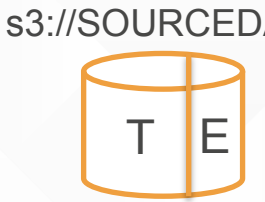
amazon web services

４つのステップ

1. 教師用/評価用データを準備
2. モデルを作成(学習、トレーニング)
3. モデルの品質評価
4. 実際の予測の実施

1.教師用/評価用データを準備
Data Sourceの作成

S3、 Amazon Redshift、 RDS上のMySQLに格納されたデータを指定し、教師データ/評価用データとして利用する



S3、Redshift、RDSが利用可能

デフォルトの設定を使うと、自動的に７割を教師データ(T)、３割を評価用データ(E)に分割して管理してくれる

2.教師データからモデルを作成

教師データを元にAmazon Machine Learningが自動的にモデルを選択してくれる。例えば予測対象のカラムが二値型であれば二項分類が自動的に選択される。

二項分類
　ロジスティック回帰

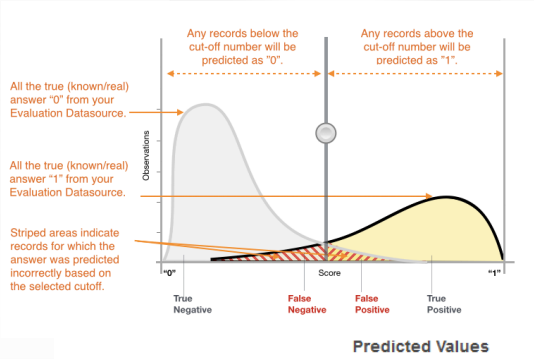
多クラス分類
　多項式ロジスティック回帰

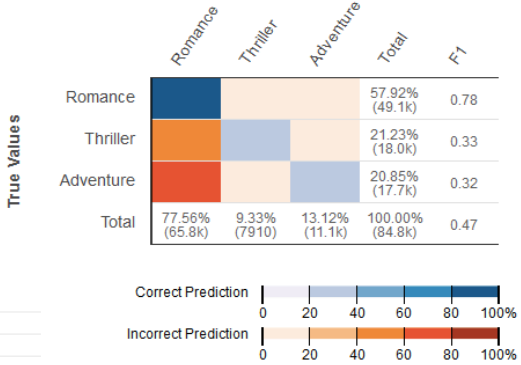
回帰分析
　線形回帰

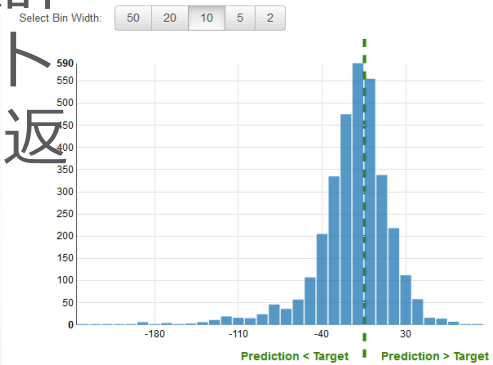
3. モデルの品質評価

作成したモデルに対して評価（評価用のデータを流してみて予測の精度を測ること）を実施する。

精度に満足できない場合、教師データのETLや量を精査し、トレーニングと品質評価を繰り返す。





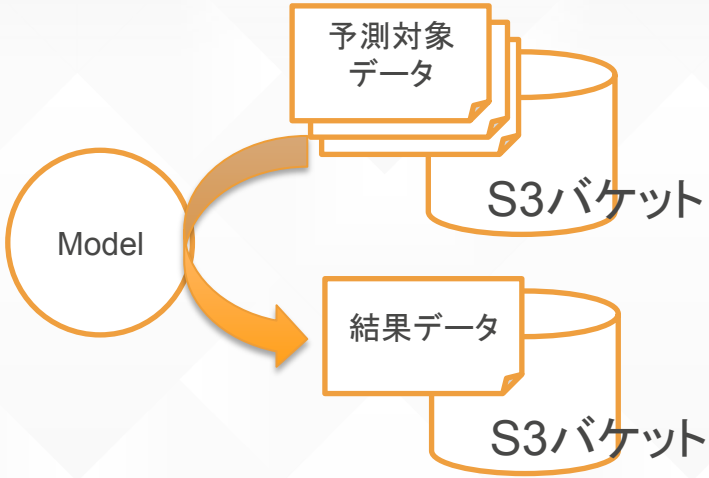


amazon web services

4.実際の予測の実施

バッチ予測

S3(S3、Redshift、RDS）等にアップロードされた予測対象データに対してまとめて予測を実施



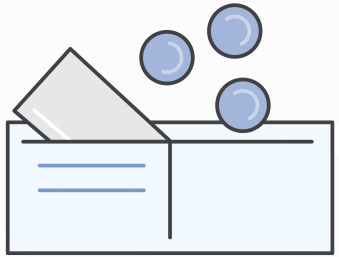
リアルタイム予測

データ１件ずつAPIを使って予測を実施する

```
ml     = Aws::MachineLearning::Client.new
record = { attr_A: ‘foo’, attr_B: ‘bar’,... }
result =  ml.predict(
    ml_model_id: MODELID,
    record: record,
    predict_endpoint: ENDPOIND
)
```

料金

使った分だけ、安価な支払い



データ分析、モデルトレーニング、評価:
$0.42/インスタンス時

バッチ予測: **$0.10/1000**

リアルタイム予測: **$0.10/1000**
+ 1時間毎のキャパシティリザベーションチャージ(モデルサイズ10MBあたり$0.001)

リージョン

amazon web services

リージョン

- 現在のところus-east-1のみ
- ただし他のリージョンのS3も利用可能

ユースケースを考えてみる

amazon web services

機械学習を活用するために

- 機械学習をうまく活用するためには、自分たちが解決したい問題を予測モデルに対してうまく落としこんでやることが非常に重要

- ここから何枚かのスライドでは、その落し込みの例について考えてみる

広告の不正クリック検出

- 教師データ
  - 実際のクリックログ。不正かそうでないかを別の手法で判断してフラグがついたもの。
- 問題の分類
  - 二項分類
- 出力
  - ログ一行ごとに不正クリックかそうでないか

デモグラ推定

- 教師データ
  - デモグラがわかっているユーザーの行動ログ
- 問題の分類
  - 多クラス分類
- 出力
  - ユーザーのカテゴライズ（たとえばF1,M1のような）されたデモグラ

デモグラに基づいたレコメンデーション

- 教師データ
  - 購入履歴にデモグラ情報がマッピングされたもの
- 問題の分類
  - 多クラス分類
- 出力
  - このユーザーはF1なので商品カテゴリ●●、みたいな出力

顔写真から特定の人物かどうかを判定する

- 教師データ
  - 顔写真をグレースケールにしたビットマップ
- 問題の分類
  - 二項分類
- 出力
  - この写真は●●さんである/違う

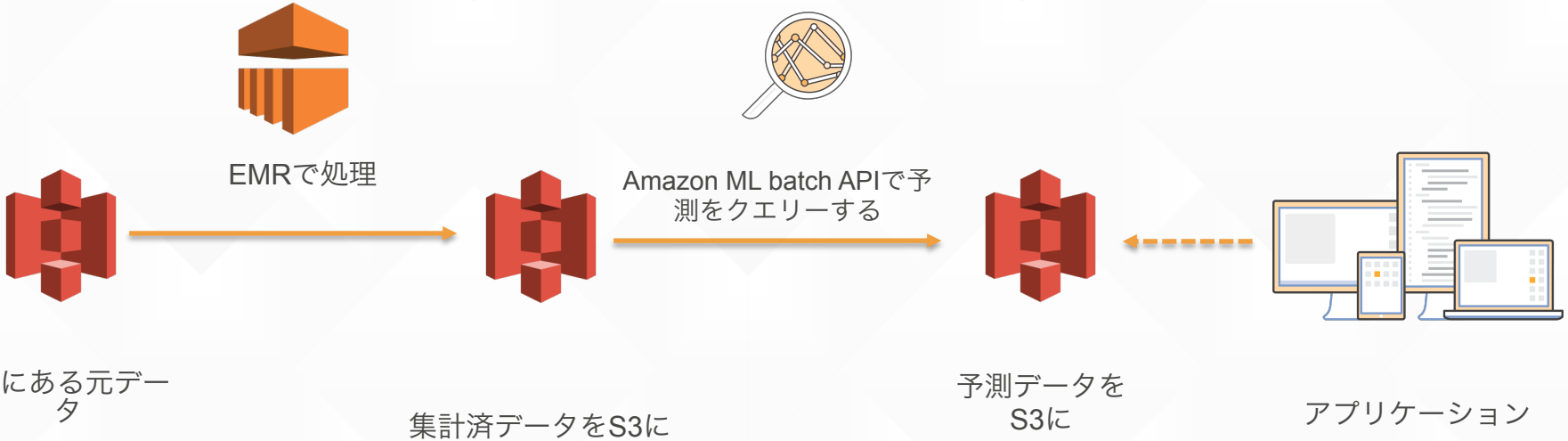
問題をどうモデルに落としこむかが非常に大事

- 教師データの量
- 教師データの形式、入力値
- どのモデルを使うのか
- 何の値を予測/判定するのか

Part3: アーキテクチャへの組み込み

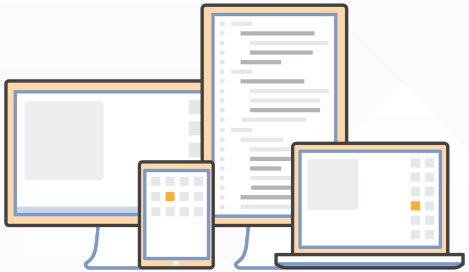
amazon web services

EMRを使用したバッチ予測



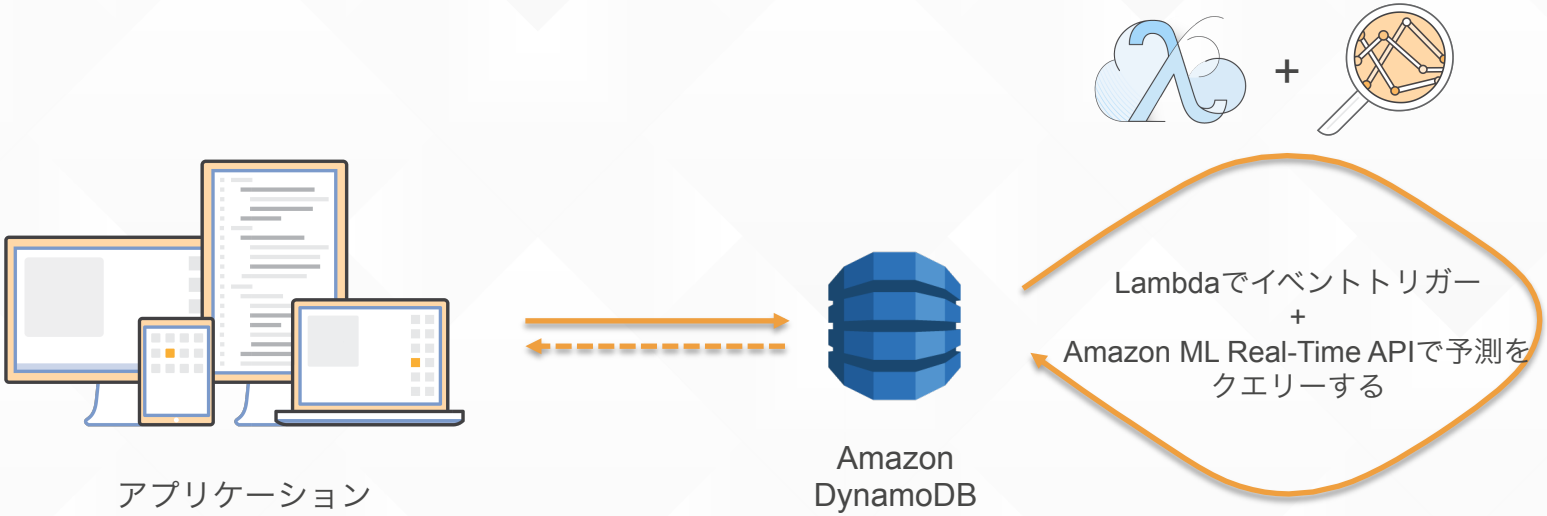
Amazon Redshiftを使ったバッチ予測

インタラクティブアプリケーション用のリアルタイム予測



アプリケーション

既存データフローに予測を追加する



さいごに

amazon web services

Amazon Machine Learningを使うと・・・

- 機械学習の導入を非常に容易にしてくれる。機械学習やそのためのシステムについての専門家がいなくても可能。
- S3やRedshiftにデータがあればすぐにでも使いはじめることができる。

溜めていたデータの評価や、その先のビジネスへの活用を簡単に始めることができる。

Thank You

amazon web services

より深い情報が知りたければ

amazon web services

OCUMENTATION HOME

mazon Machine Learning
ocumentation

ELATED LINKS
WS Glossary
etting Started with AWS
DKs & Tools
WS Documentation on Kindle
WS General Reference
WS Training
WS Case Studies
WS Whitepapers

Amazon Machine Learning Documentation

Amazon Machine Learning is a service that makes it easy for developers to build smart applications, including applications for fraud detection, demand forecasting, targeted marketing, and click prediction. The powerful algorithms of Amazon Machine Learning create machine learning (ML) models by finding patterns in your existing data. The service uses these models to process new data and generate predictions for your application.

Developer Guide
Provides a conceptual overview of Amazon Machine Learning and includes detailed instructions for using the service.
HTML | PDF

API Reference
Describes all the API operations for Amazon Machine Learning in detail. Also provides sample requests and responses for supported web service protocols.
HTML | PDF

Machine Learning Concepts
Provides an overview of the basic concepts in the field of machine learning.
HTML | PDF

http://docs.aws.amazon.com/machine-learning/latest/mlconcepts/index.html

Learning Algorithm

The learning algorithm's task is to learn the weights for the model. A learning algorithm consists of a loss function and an optimization technique. The loss is the penalty that is incurred when the estimate of the target provided by the ML model does not equal the target exactly. A loss function is a function that quantifies this penalty as a single value. An optimization technique seeks to minimize the loss. **In Amazon Machine Learning, we have three loss functions, one for each of the three types of prediction problems. The optimization technique used in Amazon Machine Learning is online Stochastic Gradient Descent (SGD).** SGD makes sequential passes over the training data, and during each pass, updates feature weights one example at a time with the aim of approaching the optimal weights that minimize the loss.

Amazon Machine Learning uses the following learning algorithms:

- For **binary classification**, Amazon Machine Learning uses **logistic regression (logistic loss function + SGD).**
- For **multiclass classification**, Amazon Machine Learning uses **multinomial logistic regression (multinomial logistic loss + SGD)**.
- For **regression**, Amazon Machine Learning uses **linear regression (squared loss function + SGD).**