

機械学習導入にあたって 理解したい 4 つのステップ



注意

本書は、情報提供の目的のみのために提供されるものです。本書の発行時点における AWS の現行製品と慣行を表したものであり、それらは予告なく変更されることがあります。お客様は本書の情報および AWS 製品の使用について独自に評価する責任を負うものとします。これらの情報は、明示または黙示を問わずいかなる保証も伴うことなく、「現状のまま」提供されるものです。本書のいかなる内容も、AWS、その関係者、サプライヤー、またはライセンサーからの保証、表明、契約的責任、条件や確約を意味するものではありません。お客様に対する AWS の責任は、AWS 契約により規定されます。本書は、AWS とお客様の間で行われるいかなる契約の一部でもなく、そのような契約の内容を変更するものでもありません。

目次

はじめに	1
人工知能 (AI)、機械学習 (ML)、深層学習 (DL) とは	2
ステップ 1: ビジネスの課題を見極める	3
ビジネス課題と ML のサイクル	3
ステップ 2: 必要なデータを準備する	5
データの選定	5
データの収集	5
データの加工	5
データの蓄積	6
ステップ 3: 機械学習基盤を手に入れる	7
サービス	7
プラットフォーム	8
フレームワークとインフラストラクチャ	8
ステップ 4: ビジネスとの融合	9
NTT ドコモ: AI タクシー	9
日本経済新聞: 完全自動決算サマリー (AI 記者)	9
エフエム和歌山: 人工知能アナウンサー	10
さあ、はじめよう	11
Amazon SageMaker	11
Amazon 深層学習 AMI	11
リソース	13
商標について	13
脚注	14

要約

アマゾン ウェブ サービス (AWS) は *Amazon* の業務に裏打ちされた IT インフラストラクチャをサービスとして提供しており、機械学習も同様です。業務の効率化や最適化に役立つ機能やインフラストラクチャを提供しています。

機械学習をビジネスに生かすには、まずビジネスの課題を特定し、次にデータを整備し、機械学習基盤を入手し、ビジネスに融合していくという流れで進めていきます。

AWS ではあらゆる規模とユースケースに対応可能な機能やサービスを提供しています。機械学習のためのフルマネージドサービス *Amazon SageMaker* や深層学習に必要なサービスがそろった *Amazon* 深層学習 AMI などがあり、すぐに導入できるのがメリットです。

本ホワイトペーパーでは、これから業務に ML 導入を検討している人に向けて、ML 導入の概要と利用可能な AWS のサービスを解説します。

はじめに

アマゾン ウェブ サービス (AWS) は 2006 年から企業向けに IT インフラストラクチャサービスを提供しています。AWS が提供するサービスは、1995 年から EC サイトとしてビジネスを始めた Amazon を支える IT インフラ技術が根底にあります。グローバルにビジネスを展開し、成長を続けてきた Amazon の業務に裏打ちされた技術であるところが AWS の特徴であり、強みです。

AWS の機械学習 (ML: *Machine Learning*) にも同じことが言えます。Amazon の業務を支える重要な技術の 1 つに ML があります。Amazon では業務で膨大なデータを処理するのと同時に、業務で生じるデータを活用してビジネスを発展させています。膨大な投資もしており、技術開発を継続しています。

Amazon のレコメンドエンジンを例に上げると、顧客の購買や検索の履歴データからおすすめ商品を提案しており、ここに ML は欠かせません。ほかにも在庫管理、売上予測、価格設定など Amazon では業務の多くで ML を使用して効率化や最適化を実現しています。さらに、スマートスピーカー Amazon Echo の中枢となる音声サービス Alexa、ドローンで配達する Prime Air、レジのない店舗 Amazon GO など、Amazon では至るところに ML 技術を活用した新サービスを提供しています。

現在、多くの企業やベンダーが ML 技術を磨きあげています。例えば画像から顔を識別する、文章を各地の言語に翻訳するなど、自社が提供するサービスと保有するデータにより、各社が得意とする ML の分野は異なります。AWS の ML は Amazon の業務で培われているところが大きな特徴です。それゆえに業務に ML を導入しようとするなら AWS は最適な選択肢となりえるでしょう。

ML で AWS を選ぶ大きな理由の 1 つにクラウドをベースとしたプラットフォームであることが挙げられます。ML では大量のデータを保有し、大量のコンピューティングリソースが必要になります。AWS なら同じクラウドのプラットフォームで ML に関係する全ての処理を統合することができるため、効率のよい ML の導入が可能になります。



人工知能 (AI)、機械学習 (ML)、深層学習 (DL) とは

人間の知能を人工的な機械で再現しようとする試みから「人工知能 (AI)」の研究が始まりました。1950 年代からはじまったコンピュータサイエンスの 1 分野です。1980 年代ごろに ML が登場し、さらに 2010 年代に深層学習 (DL: Deep Learning) が登場し、近年発展が加速しています。

ML や DL では、機械に大量のデータを与え、正しい判断ができるように学習させ、特定の事象についてのモデル (知見) を獲得させます。このモデルを元に判断や予測を行います。ML では機械が正しく判断できるように、人間が見本や特徴を示します。例えば不良品の異常検出、不正取引や不正アクセスの不正検出などで使われています。

DL は ML から発展した手法の 1 つで、機械が自らデータから特徴を抽出できるようになっています。画像や動画の分類、音声認識、自然言語の理解などで日々発展している分野です。

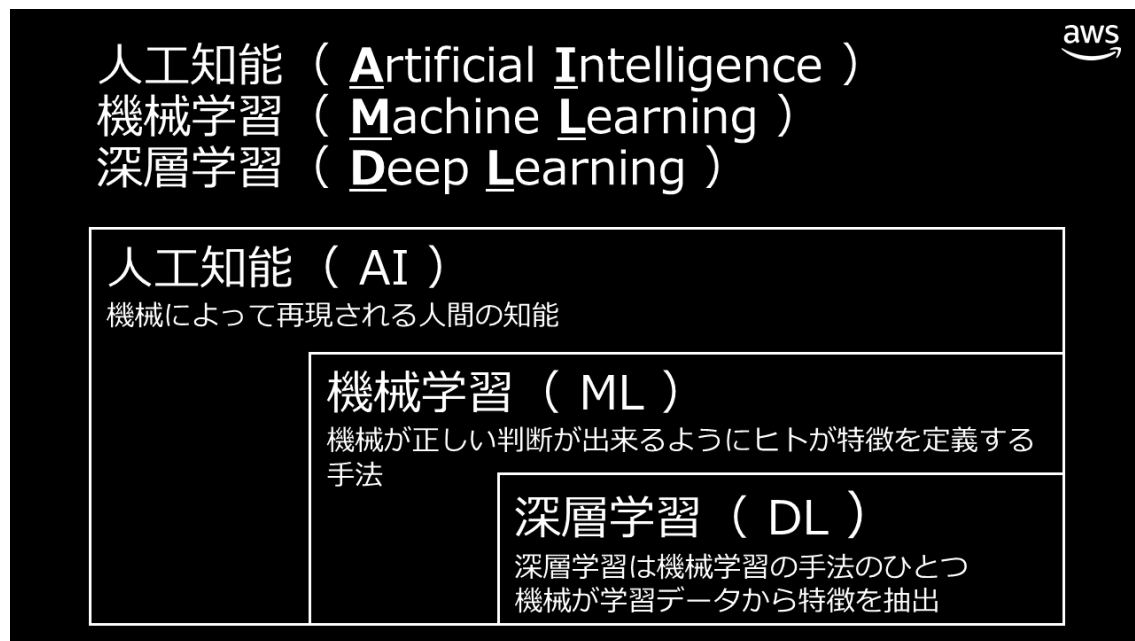


図 1: 人工知能、機械学習、深層学習の図

ビジネスソリューションから見ると、購買履歴からクロスセル商材を決定して売上向上につなげる、販売実績から将来の需要を予測して自動発注の仕組みを構築する、機器のセンサーから故障予測して生産ラインの安定化につなげるなどの目的に対して ML のスコアリングや異常検知が有効です。

ステップ 1: ビジネスの課題を見極める

「ビジネスに ML を導入するなんて敷居が高い」「どこから始めたらいいのかわからない」など、漠然とした疑問や不安を抱えている人は少なくありません。しかし目的が漠然としたままだと、すべきことが見えてきませんし、ML 導入が成功したかどうか不明瞭になります。現在、ビジネスでどのような課題を抱えており、どの課題に対して ML を活用するのかを見極めることが最初のステップとして重要です。

実存するビジネスを思い浮かべてみてください。把握できればビジネスにメリットがある予測にはどのようなものがあるでしょうか。例えばリピーターになりそうな顧客、逆に解約を検討していそうな顧客の特徴で考えてみましょう。もしこれが把握できたら、顧客に適切な商品や契約の提案ができるかもしれません。それだけではなく、顧客維持費用の最適化や適切な経営判断につながれるかもしれません。把握できればビジネスに有益な予測となります。

今、把握したい何かに対して既にビジネスインテリジェンス (BI) を用いているなら、次のステップとして ML 導入に進みやすいと言えます。BI は主に過去の把握や可視化に重点が置かれているものの、ML で BI を補完することで将来の予測へとつなげることができるからです。

ビジネス課題と ML のサイクル

ML はデータ活用という点では BI と共通しており、活用のサイクルもまた BI と似ています。ビジネス課題を解決するために ML があり、ML はデータ管理、モデル評価、アプリケーション開発、予測をサイクルで繰り返し、予測の精度を高めていきます。

データ管理では、データエンジニアやデータベース管理者が ML に使うデータを管理します。ML に適したデータ構造や属性を持ち、正確性や鮮度も重要となります。

モデル評価では、データ分析担当のデータサイエンティストがフレームワークを用いてモデルを設計し、結果を評価していきます。

アプリケーション開発では、開発担当者が ML で得られる結果をアプリケーションに組み込んでいきます。



予測では、フロントエンド担当者が分析結果や予測を展開していきます。

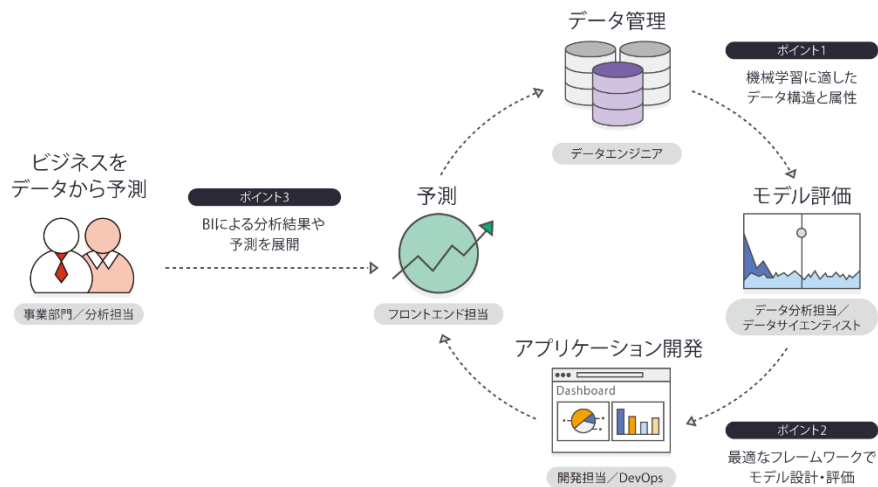


図 2: 企業への機械学習の導入

AWS ではそれぞれの段階で下記のようなサービスが利用可能です。

データ管理: Amazon Simple Storage Service (S3), Amazon Redshift など

学習およびモデル評価: Amazon Machine Learning (Amazon ML), Amazon SageMaker, AWS 深層学習 AMI, 各種フレームワーク (TensorFlow, MXNet, Caffe など)

アプリケーション開発: AWS CodeBuild, AWS CodeCommit, AWS CodeDeploy など

ステップ 2: 必要なデータを準備する

ML はデータから学習してモデルを導き出すのでデータが不可欠です。ML を料理に例えたらデータは材料のようなものです。正確なモデルや結果を導き出すためには、データが適切である必要があります。いかに必要かつ適切なデータを準備できるかどうか ML の成否を決めると言っても過言ではありません。

データの選定

データを過不足なく ML に与えられるように、必要なデータを選定します。データの品質や鮮度も重要です。例えば不良品検知の学習をさせるとき、商品と無関係なデータが混じっていると ML は間違った学習をしてしまいかねません。

データの収集

必要なデータは散在しているかもしれません。あるデータは企業システムのデータベース、ほかのデータは社外で公開されているオープンデータやインターネットの SNS など、分散していたら ML が処理する場所に集める必要があります。必要なデータを適切なタイミングで収集できるようにします。

AWS へのデータベース移行には *AWS Database Migration Service*、ストリーミングデータのリアルタイム収集には *Amazon Kinesis*、オンプレにあるデータをクラウドに転送するなら *AWS Snowball* などが利用可能です。

データの加工

データを ML 処理にかける時、ML に適したデータ構造や形式に合わせて加工が必要になる場合があります。

AWS では、データの抽出、変換、ロード (ETL) をマネージド型サービスで提供する *AWS Glue* がご利用いただけます。



データの蓄積

データには蓄積しておく場所が必要になります。いわゆるデータレイクとなる場所です。AWS ならスケーラブルなオブジェクトストレージの *Amazon Simple Storage Service (S3)*、管理された *Hadoop* フレームワーク *Amazon EMR* などがご利用いただけます。あるいは *BI* または *ML* のプラットフォームにデータを蓄積します。

ステップ 3: 機械学習基盤を手に入れる

ML を稼働させるにはコンピューティングリソースやフレームワークなどで構成された機械学習基盤が必要です。ML を料理に例えたら機械学習基盤はキッチンのようなものです。人工知能や ML に関連する機械学習基盤は主に 4 つのレイヤーに分けることができます。それがサービス、プラットフォーム、フレームワーク、インフラストラクチャです。

機械学習基盤の 4 つのレイヤーを AWS が提供するサービススタックに照らし合わせると下図（図 3）のようになります。AWS では規模に応じて拡張が自在で、各種マネージドサービスもそろっているため機械学習基盤の構築から運用の手間を省力化し、すばやく ML を導入できます。

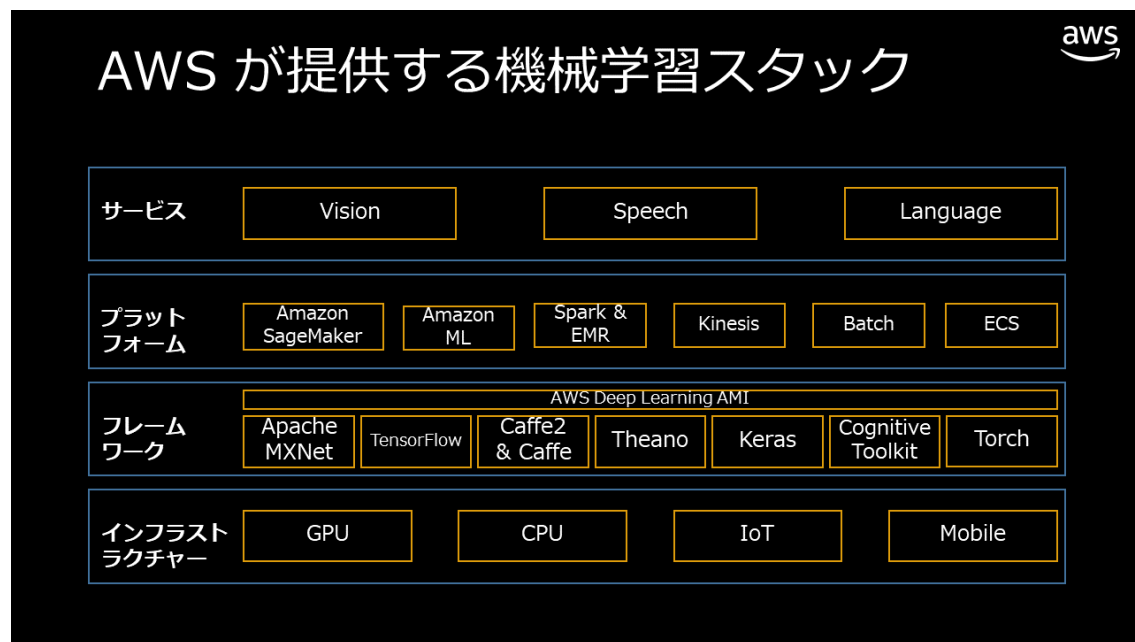


図 3: AWS が提供する機械学習スタック

サービス

サービスのレイヤーで提供しているのは API で連携できる人工知能サービスです。画像や動画分析なら *Amazon Rekognition Image* や *Amazon Rekognition Video*、チャットボットなら *Amazon Lex*、翻訳なら *Amazon*



Translate、テキストから音声への変換なら *Amazon Polly*、文字起こしのような音声からテキストへの変換なら *Amazon Transcribe*、テキスト内でインサイトや関係性を検出する *Amazon Comprehend* などの機能が API で利用できます。AWS によって学習済みなので API コールするだけで使えます。対応言語については最新情報をホームページでご確認ください。

プラットフォーム

プラットフォームのレイヤーは ML 処理そのものが稼働する場所です。AWS では、ML を簡単に利用できる *Amazon ML* を提供しています。ML モデルの作成プロセスをウィザードで提供するので、複雑な ML アルゴリズムやテクノロジーに詳しくなくても ML を始めることができます。*Amazon* 社内のデータサイエンティストコミュニティで使用された実績があり、高度にスケーラブルであることも特徴です。

2017 年 11 月には AWS は *Amazon SageMaker* を発表しました。これは ML モデルの構築、トレーニング、デプロイがセットになっているフルマネージドサービスです。データサイエンティストや開発者はあらゆる規模で、迅速かつ簡単に機械学習基盤を構築し、トレーニングやデプロイができます。

フレームワークとインフラストラクチャ

フレームワークとインフラストラクチャのレイヤーは ML が必要とする環境です。AWS では *Apache MXNet* や *TensorFlow* などの主要な ML のフレームワークをサポートしています。AWS は、フレームワークと *Amazon EC2* がセットになった AWS 深層学習 AMI があります。素早く環境を構築することができます。

AWS ではユースケースに適したコンピューティングリソースを選べるのも特徴です。2017 年 10 月に登場した *Amazon EC2 P3* インスタンスでは *NVIDIA Tesla V100* を最大 8 基搭載できて、従来の *P2* インスタンスよりも最大 14 倍高い性能を発揮することができます。

ステップ 4: ビジネスとの融合

AWS の ML や DL は既存のビジネス課題を解決するだけでなく、ビジネスに新たなサービスとして融合させる事例が出てきています。下記事例のほかに、海外でも数多くの ML 活用事例があります。映像ストリーミングサービスの Netflix ではレコメンドエンジン、画像共有サービス Pinterest では画像検索、オンライン不動産データベースの Zillow では不動産価格推定、医療機関情報提供の Zocdoc では医療機関の検索や予約を実現しています。

AWS Summit Tokyo や AWS Solution Days など、AWS 主催イベントで紹介した ML を活用した国内の事例をいくつか紹介します。

NTT ドコモ: AI タクシー

AI タクシーとは、AI で将来のタクシー需要予測を配信するサービスです。タクシーの乗車実績や天気予報に加えて、NTT ドコモが基地局を通じて集めた携帯電話利用者の位置情報から人の動きをビッグデータ分析することで、エリアごとのタクシーの需要を予測して配信します。

タクシー乗車台数の予測は、営業区域内を 500m 四方で区切ったエリアごとに、現在から 30 分後まで 10 分ごとに配車システムへ提供します。実際にタクシードライバーが見る画面では、地図上に予測の数値が大きいエリアは赤、少なければ青など、予測に応じた色分けがされています。

タクシードライバーが需要予測を把握できることで、需要と供給のマッチングが向上し、タクシー運行が効率化します。タクシー会社は供給を最適化し、乗客はタクシーをつかまえやすくなるというメリットが生まれます。

日本経済新聞: 完全自動決算サマリー (AI 記者)

日本経済新聞では、上場企業が開示する決算発表資料から人工知能が決算サマリー記事を作成しています。2015 年から東京大学松尾研究室と共同研究を開始し、言語理解研究所 (ILU) の技術も用いています。

決算発表資料をもとに業績、要因、見通しの 3 パラグラフを作成し、人工知能技術によって文章作成からインターネット公開に至るまでの作業を自動化しています。これまでは年 4 回の決算発表で記者 1 人につき、50~70 社を



担当していたところ、決算発表がピークを迎える時期でも短時間で大量の記事を生成できます。証券取引所から決算短信が発表されると、1~2 分で処理が完了します。現時点では流暢さや創造性においては課題があるものの、正確性や処理能力が強みになります。

決算サマリーのほか、市況速報やプレスリリースなど定型的な業務を AI に任せることで、人間の記者はより付加価値の高い業務に集中できるようになります。

エフエム和歌山: 人工知能アナウンサー

エフエム和歌山では和歌山県和歌山市を中心に約 50 万人を対象にコミュニティ FM 放送「バナナエフエム」を提供しています。2017 年 7 月から「人工知能アナウンサー・ナナコ」がニュースや天気予報を読み上げています。

活用しているのはテキストから音声変換を行う *Amazon Polly* とアプリケーションのセキュリティを強化する *Amazon Cognito* です。*Amazon Polly* にかかるコストは年間 400~800 円程度という安さが大きな魅力です。

安定してアナウンサーを確保することができない時間帯や状況でも、安定した読み上げのクオリティで無人放送を実現しています。2017 年 9 月に台風 18 号が通過した時には和歌山市内で最大 4400 世帯が停電するなか、ナナコは情報を随時更新しながら天気と災害情報をアナウンスし続けました。災害時でも適切に情報を配信できるようにすることで、1 人でも多くの命を救えるようにと期待されています。

さあ、はじめよう

ML をはじめるには、これまで述べてきたように、データを準備し、機械学習基盤となる環境を準備する必要があります。環境準備には、処理を行うサーバー、データを蓄積するストレージ、ML 処理に必要なソフトウェアを OS からアプリケーションまで一通り選定してインストールしなくてはなりません。クラウドサービスを用いて、ハードウェアや OS などのインフラストラクチャをある程度準備できたとしても、データレイク構築やフレームワーク導入など、準備作業は多数あります。

しかし AWS ならすぐに ML をはじめられます。AI や ML に関連するサービスが一通りそろっており、必要に応じて取捨選択できて連携も容易です。ML に利用するためのデータの準備が難しい場合でも、AWS によって学習済みのサービス *Amazon Lex*、*Amazon Polly*、*Amazon Rekognition* を使えば、すぐに ML を導入することができます。

さらに本格的に ML を導入する場合には、ML に必要な機能がセットとなり、保守が不要なマネージドサービスという形で提供されている *Amazon SageMaker* や DL に適した *Amazon 深層学習 AMI* は業務に AI や ML を導入するうえで最適なソリューションとなりえます。

Amazon SageMaker

Amazon SageMaker は機械学習モデルの構築、トレーニング、デプロイのモジュールをマネージド型サービスで提供しています。構築モジュールにはデータ操作のインスタンス、高いパフォーマンスのアルゴリズムの組み込み、幅広いフレームワークなど環境構築に必要なものがそろっています。トレーニングを開始すると分散型コンピュータクラスターのセットアップ、チューニング、クラスター解体までを自動で行います。デプロイモジュールではワンクリックでモデルをデプロイするほか、モデルの A/B テストもできてモデルの比較も可能です。

Amazon 深層学習 AMI

AWS 深層学習 AMI は DL のためのインフラストラクチャとツールを提供しています。*Amazon EC2* インスタンスに一般的な深層学習フレームワークがインストールされています。高度なカスタム AI モデルのトレーニング、新しい



アルゴリズムの実験、新しいスキルや技術の学習が行えます。尚、Amazon EC2 の P2 または P3 インスタンスには必要なドライバーがインストールされているため、短時間で処理を完了させることができます。

AWS では、AWS の使用の開始を支援するために無料利用枠を導入しています。無料利用枠を利用して、今すぐ使用を開始しましょう！

リソース

リソース	説明
人工知能 (AI) とは¹	人工知能や機械学習の概要を解説しています。
Amazon SageMaker²	機械学習モデルの構築、トレーニング、デプロイをフルマネージドサービスで提供するサービスです。
Amazon 深層学習 AMI³	深層学習アプリケーションを構築するために事前に構成された環境です。
Amazon ML Solutions Lab⁴	機械学習のスキルや方法を伝える取り組みです。
AWS 無料利用枠⁵	AWS のプラットフォーム、製品、およびサービスを無料で体験できます。

ドキュメントの改訂

改訂日	説明
2018 年 3 月	初版発行

商標について

本ドキュメントに記載されている会社名、システム名、製品名は一般に各社の登録商標または商標です。なお、本文および図表中では、「™」、「®」は明記しておりません。

-
- ¹ <https://aws.amazon.com/jp/machine-learning/what-is-ai/>
 - ² <https://aws.amazon.com/jp/sagemaker/>
 - ³ <https://aws.amazon.com/jp/machine-learning/amis/>
 - ⁴ <https://aws.amazon.com/jp/ml-solutions-lab/>
 - ⁵ <https://aws.amazon.com/jp/free/>