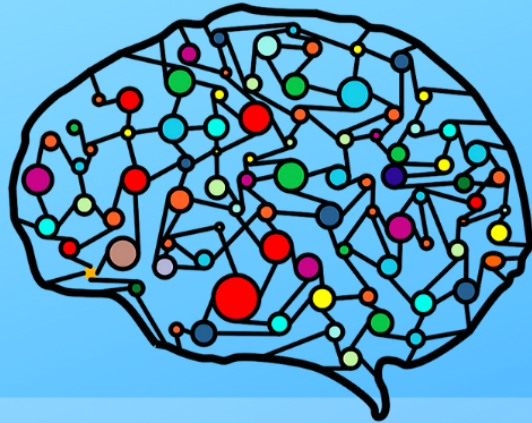


機械学習をどこよりもわかりやすく解説！ 教師ありなし学習・強化学習だけでなく5つのアルゴリズムも完全理解！

最終更新日：2019年11月27日



機械学習をとことん解説！ 機械学習マスターに俺はなる！

AINOW編集部作成

こんにちはAINOWインターンのさとしです。今回の記事ではAIやディープラーニングと混同されがちな機械学習について、それらの関係性・違いを理解できるように細かく説明します。

また、機械学習を知る上で必要不可欠な用語（教師あり学習や教師なし学習、各アルゴリズムなど）も頭の中でごちゃごちゃになってしまっていると思いますが、この記事を読むことで、そのような用語を網羅的に、且つきちんと整理・理解できるようにわかりやすく説明していきます。

目次 [\[非表示\]](#)

[機械学習とは](#)

[機械学習・AI・ディープラーニングはどのような関係？](#)

[ディープラーニングと機械学習は何が違うの？](#)

[Pythonと機械学習の関係は？](#)

機械学習は3つに分けられる！

教師あり学習

教師なし学習

強化学習

機械学習で使われるアルゴリズムを紹介

分類

回帰

クラスタリング

次元削減

異常検知

最後に

機械学習とは

機械学習とはAIの1つの要素技術です。

多くの企業で取り組むことができる技術の1つでしょう。機械学習について理解するのに必要なことは3つあり、以下ようになります。

データからルールやパターンを発見する方法である

識別と予測が主な使用用途である

分析の精度は100%ではないが、従来の手法より精度をあげられる可能性は高い

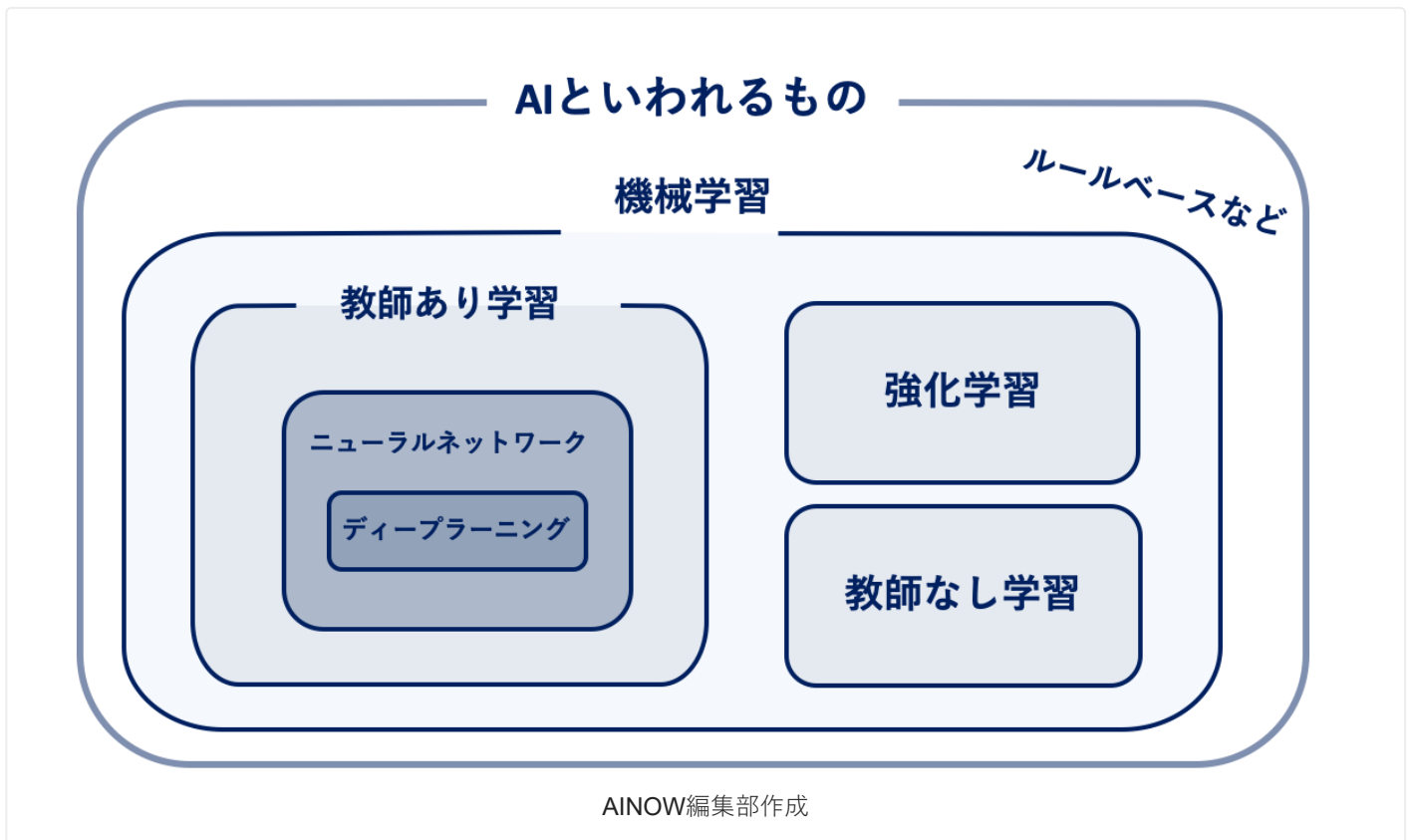
機械学習・AI・ディープラーニングはどのような関係？

世間を賑わしているAIや機械学習とディープラーニングの違いをしっかりと理解している人は多くはありません。

「AIやディープラーニングといった最新のテクノロジーを駆使し、」のような宣伝文句をしばしば見かけますが、多くがその意味を混同したまま使われているのが現状です。

3つの技術の関係性と違いを理解しておくことで、AIを導入する上で認識を合わせることが可能です。

以下の画像をご覧ください。



上記の画像から分かるように、機械学習はAIという概念の一部であって「AI＝機械学習」ではありません。大前提として「AI > 機械学習 > [ディープラーニング](#)」のイメージは持っておきましょう。

ディープラーニングと機械学習は何が違うの？

ディープラーニングは機械学習のうちの一技術です。

機械学習とは「機械に大量のデータからパターンやルールを発見させ、それをさまざまな物事に利用することで判別や予測をする技術」のことです。

その際に、データのどの部分が結果に影響を及ぼしているのか（**特徴量**という）を人間が判断し、調整することで予測や認識の精度をあげています。

一方、ディープラーニングはパターンやルールの発見に加えて、**特徴量の設定、学習も自動的に行う**のが特徴です。人間が判断する必要がないのが画期的です。

特徴量の設計を行う必要がないため、人の認識・判断では限界があった画像認識・翻訳・自動運転といった技術が飛躍的に上がったのはディープラーニングで人間が見つけられない特徴を学習できるようになったおかげとも言えます。

Pythonと機械学習の関係は？

Pythonは機械学習で最も使われているプログラミング言語です。

[IBMの調査](#)によると、「機械学習における最も人気なプログラミング言語」という項目でPythonは2016年にJavaを超えて1位になるほど機械学習における重要なプログラミング言語となっています。

Pythonには以下の3つの特徴があります。

シンプルさ

信頼性

使い勝手の良さ

この3点が際立っていることが利用される理由でしょう。

Pythonは数学をあまり知らない初等教育の子どもでも取り組めるほどシンプルさがあり、「Google」や「Facebook」などでも使われている信頼度の高いプログラミングです。

Djangoや**bottle**といったPythonを支えるフレームワークやライブラリーが非常に充実しており、使い勝手が良いのも多くの人に利用される要因です。

以下の記事ではPythonの勉強方法の勉強方法をまとめているのでぜひ参考にしてみてください。

AI専門ニュースメディア AINOW **5 users**

AI開発に不可欠なPython！独学からスクールまで勉強法をまとめてみた

<https://ainow.ai/2019/09/20/175543/>

現在、AIの国内市場規模は3000億円にも登ると言われています。AI技術の中でも特に注目されているディープラーニングなどの機械学習技術の分野で最もよく使われているのがPython（パイソン）です。この記事はなぜPythonが選ばれるのかといった話から、Pythonを学ぶた...

機械学習は3つに分けられる！

機械学習はデータのタイプや状況によって教師あり学習・教師なし学習・強化学習の3つに大きく分けられます。

教師あり学習

教師あり学習は「正しいデータ」という名の教師がいて、そこに不明なデータを持ち寄った際には正解を教えてくれるというイメージです。正解となる膨大なデータを学習することで、新しいデータに対しても対応が可能です。



AINOW編集部作成

教師あり学習は「学習」と「認識・予測」の2段階のプロセスに別れています。まず、正解がわかっているデータ（入力という）を元に、ルールやパターン（出力）を学習します。そのあとに、インプットされた「正解がわからないデータ」に対してルールやパターンを元に認識や予測をするという流れになっています。

回帰や**分類**が代表的な手法になります。

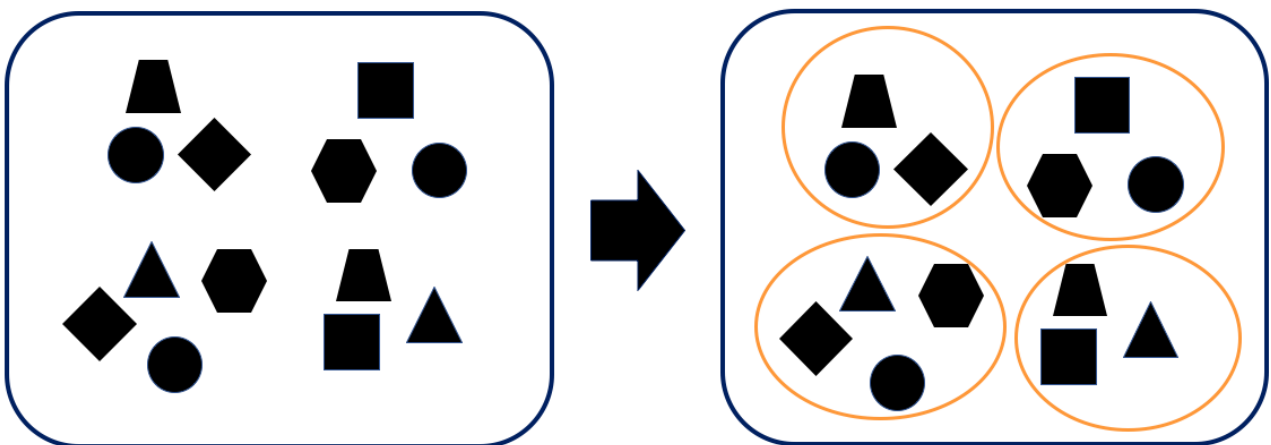
ちなみにニューラルネットワークやディープラーニングはこの教師あり学習の発展形になります。

教師なし学習

教師あり学習と違い、教師なし学習は膨大な教師データを学習しません。そこで、教師なし学習では主にデータそのものが持つ構造・特徴を分析し、**グループ分け**や**データの簡略化**といったことをします。

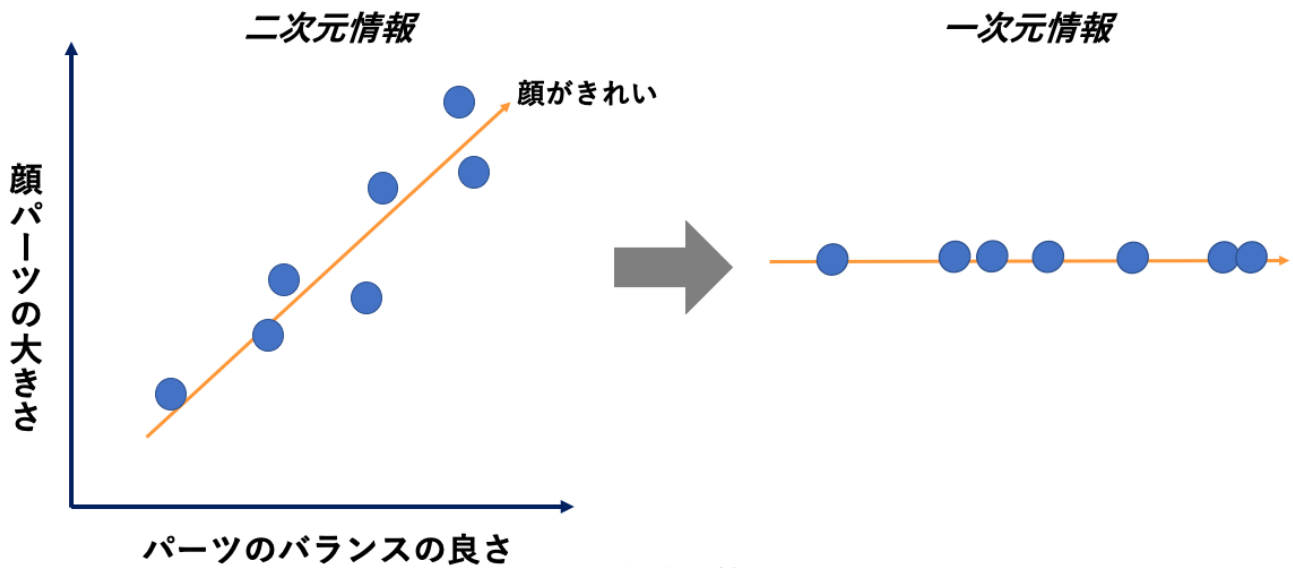
教師なし学習のうち、**クラスタリング**と**次元削減**が代表的な手法になります。以下の画像がイメージです。

クラスタリングのイメージ



AINOW編集部作成

次元削減のイメージ



※あくまで一例です

AINOW編集部作成

強化学習

強化学習とは簡単に言うと「機械がとる行動の戦略（指針）を強化（改善）する仕組み」を学ぶ手法になります。

一連の行動をとった結果ごとに報酬を設定し、その「報酬が最大化」するように機械が試行錯誤し、学習することで精度をあげていきます。

イメージとしては自転車に乗れるようになるプロセスに似ています。「乗ってみる」→「倒れる」→「乗り方を変える（強化学習の”戦略”に相当）」→「少し乗れる（強化学習の”報酬”に相当）」→「さらに乗り方を変えて徐々に乗れるようになる」の繰り返しの様な感じです。

今では強化学習とディープラーニングを組み合わせた**深層強化学習**が主流になっており、囲碁AI「AlphaGO」が世界チャンピオンを倒すまでに成長した一因となっています。

機械学習で使われるアルゴリズムを紹介

機械学習ではアルゴリズムを使用してデータを分析します。世の中には多くのアルゴリズムが存在していますが、データや状況によって使うアルゴリズムも異なってくるのでそれぞれの大まかな特徴を掴みましょう。ちなみに機械学習のアルゴリズムを選定には[scikit-learnのチートシート](#)や[マイクロソフトのチートシート](#)を参考にするといいです。

機械学習のアルゴリズムは大きく分けると以下の5つのカテゴリーに分けられます。

分類（教師あり学習）

回帰（教師あり学習）

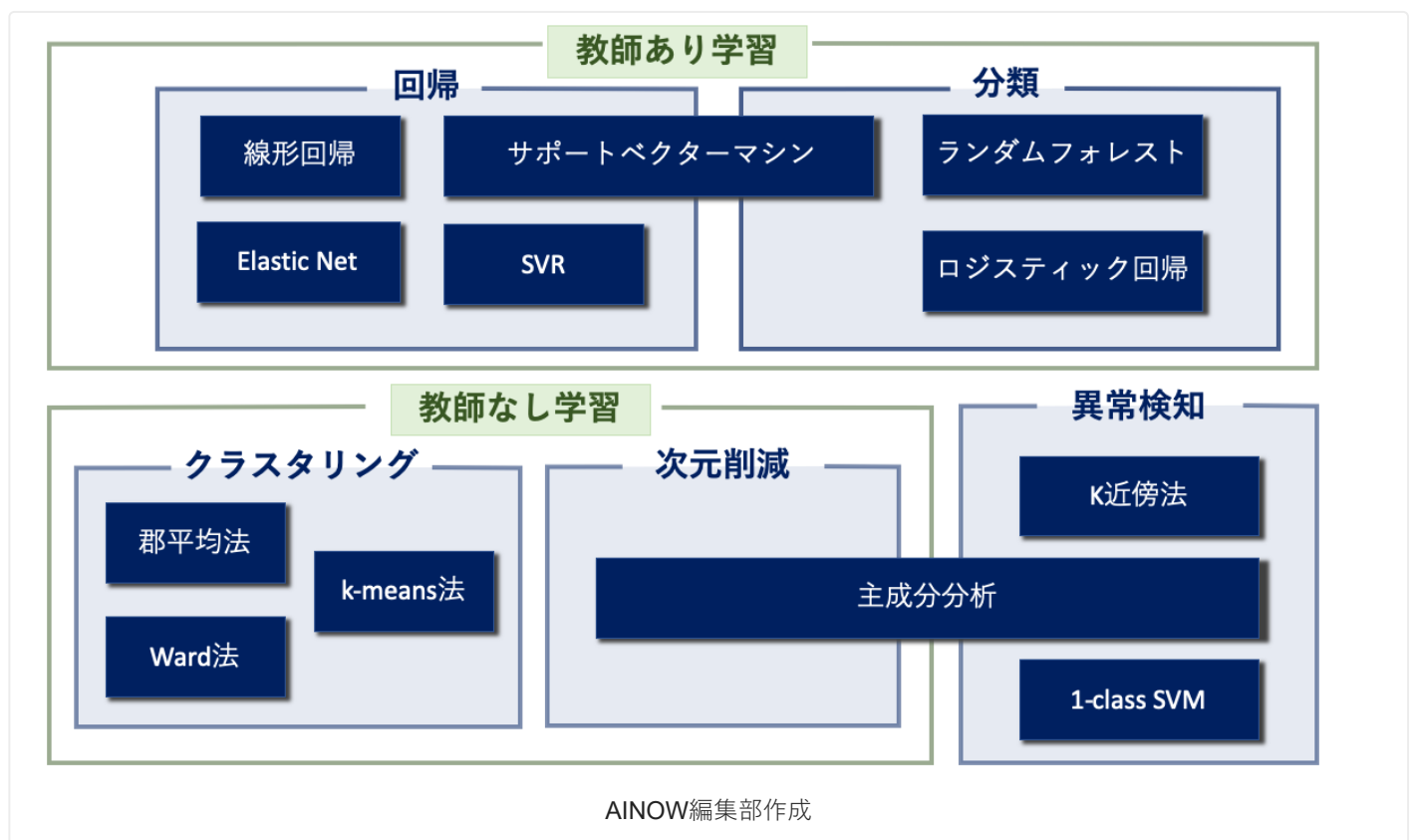
クラスタリング（教師なし学習）

次元削減（教師なし学習）

異常検知

これから紹介するアルゴリズムと上記で紹介した**教師あり学習**などを1つにまとめると、以下のような画像になります。

ここで紹介しているのはあくまで一例で、例えば[ランダムフォレスト](#)は回帰やクラスタリングでも使える場合もあるなど、この仕分けは絶対的なものではないことは心に留めておいてください。

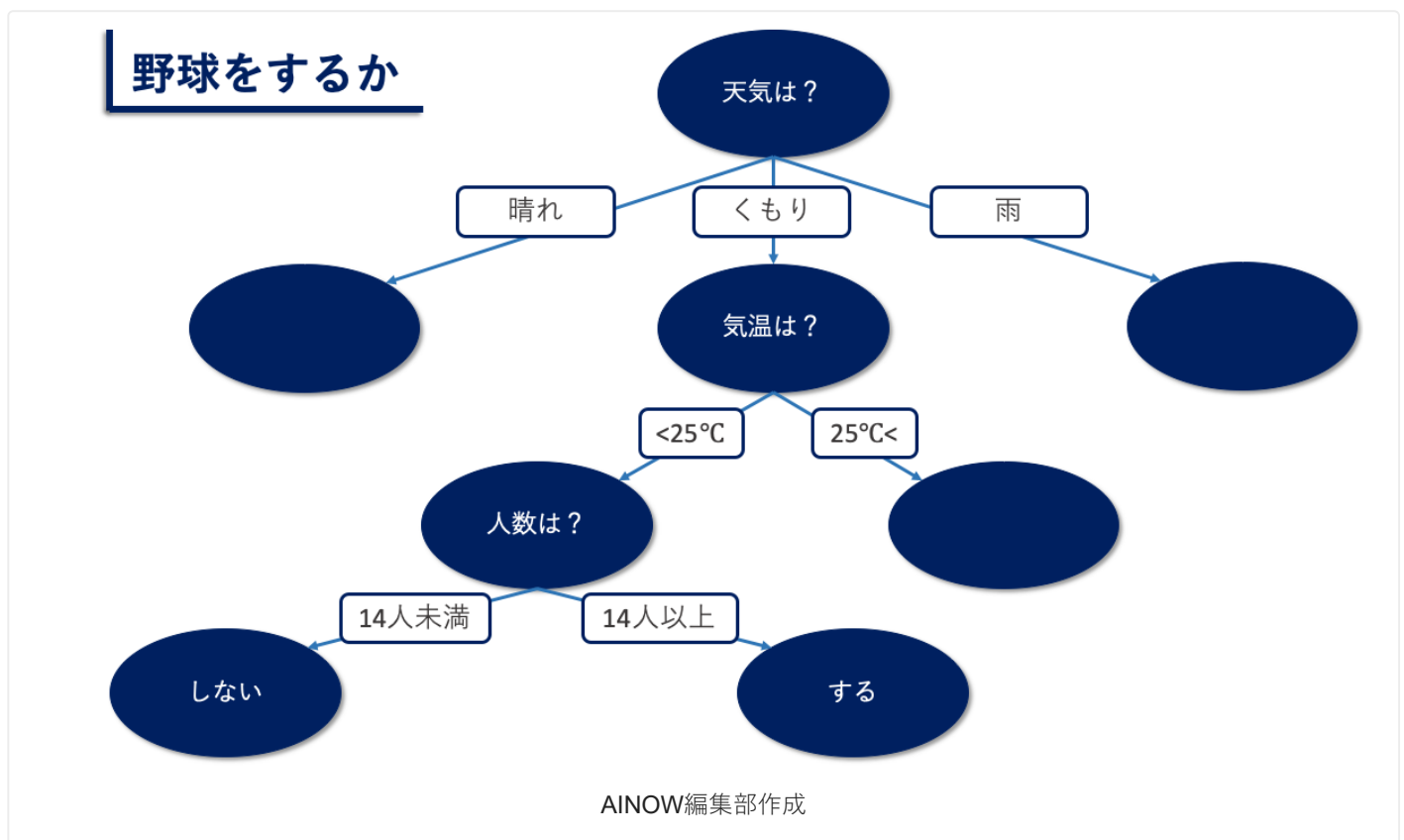


分類

分類とは機械学習のうちの「教師あり学習」の1つで「犬」や「猫」などのラベルの情報や「購入」「非購入」といったカテゴリーの情報などを分類・予測する場合の学習方法です。

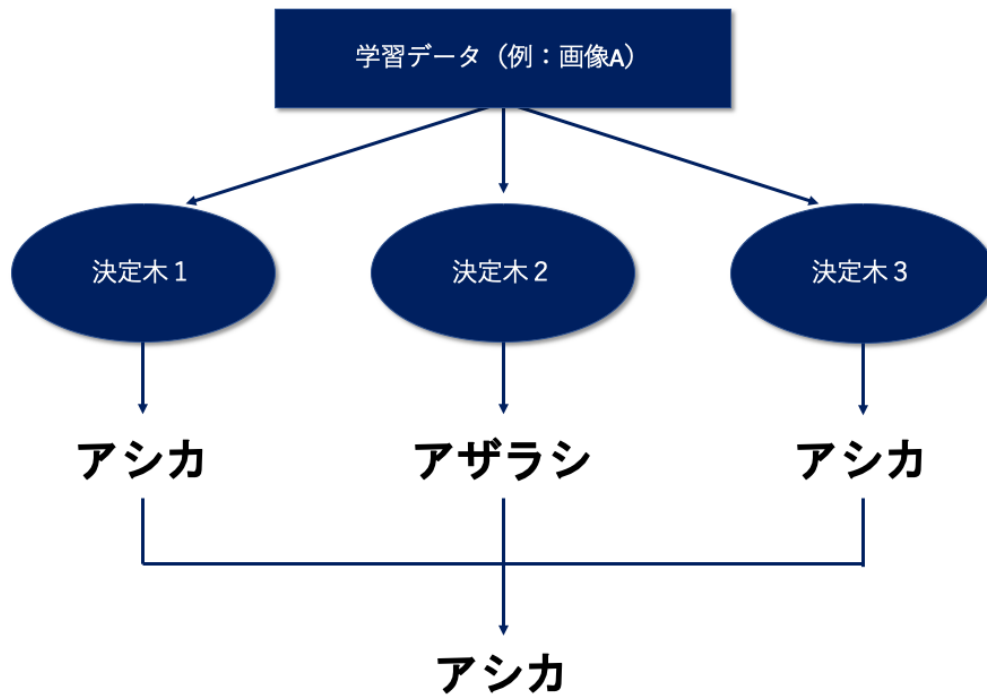
ランダムフォレスト

代表的な手法の1つにランダムフォレストという決定木を利用した手法があります。決定木とはリスクマネジメントをはじめ、何かしらの決定をする際に利用され樹木形をした以下のようなグラフのことです。



ランダムフォレストは複数の決定木で出た結果を用いて多数決をとって最終的な結果を出力する仕組みをとっています。

この手法を使うことで、精度の悪い決定木があったとしても全体としての精度を守ることができる仕組みになっています。以下の画像がそのイメージです。



AINOW編集部作成

ロジスティック回帰

線形回帰という手法を応用したもので、名前に回帰とついているのですが、実際は分類問題に使用される手法です。つまり目的変数が「Yes (= 1)」か「No (= 0)」のどちらかに分類できるような問題を扱うということです。

この手法では量的変数から質的変数を推測するのですが、変数の値を予測するのではなく、目的変数が"1"になる確率を予測します。

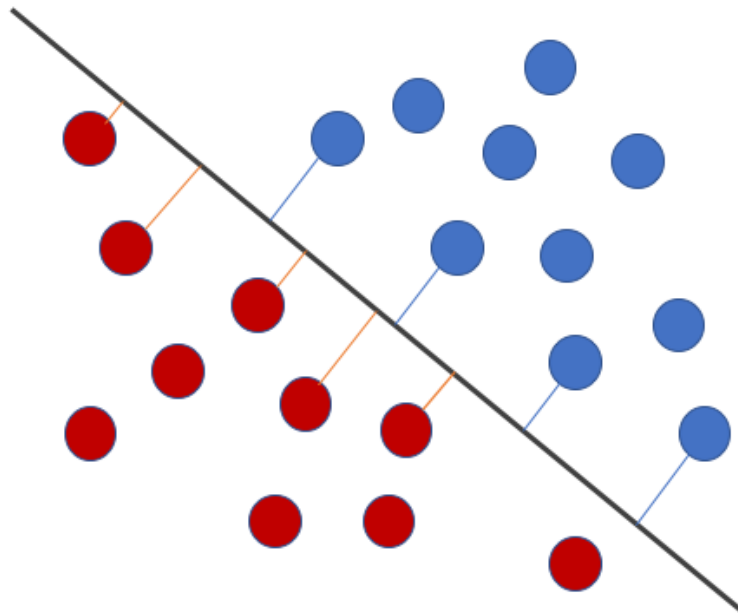
そのため重回帰分析などで1以上の数値が出てわかりにくかった場合でも、ロジスティック分析では必ず0から1の範囲で確率がわかるので分析結果を見たときに理解しやすくなるのです。

サポートベクターマシン

サポートベクターマシンは**分類・回帰**の両方に利用可能な教師あり学習です。非線形な識別をするための実装が容易なため、識別能力が高く、精度が良い分析結果を出してくれることから人気の高いアルゴリズムになっています。

サポートベクターマシンはデータ分布を複数のクラスに明確に分けるために境界を引く手法で、イメージは以下の画像です。

それぞれの点との距離の総和が最大になるような線を引く（マージン最大化という）



AINOW編集部作成

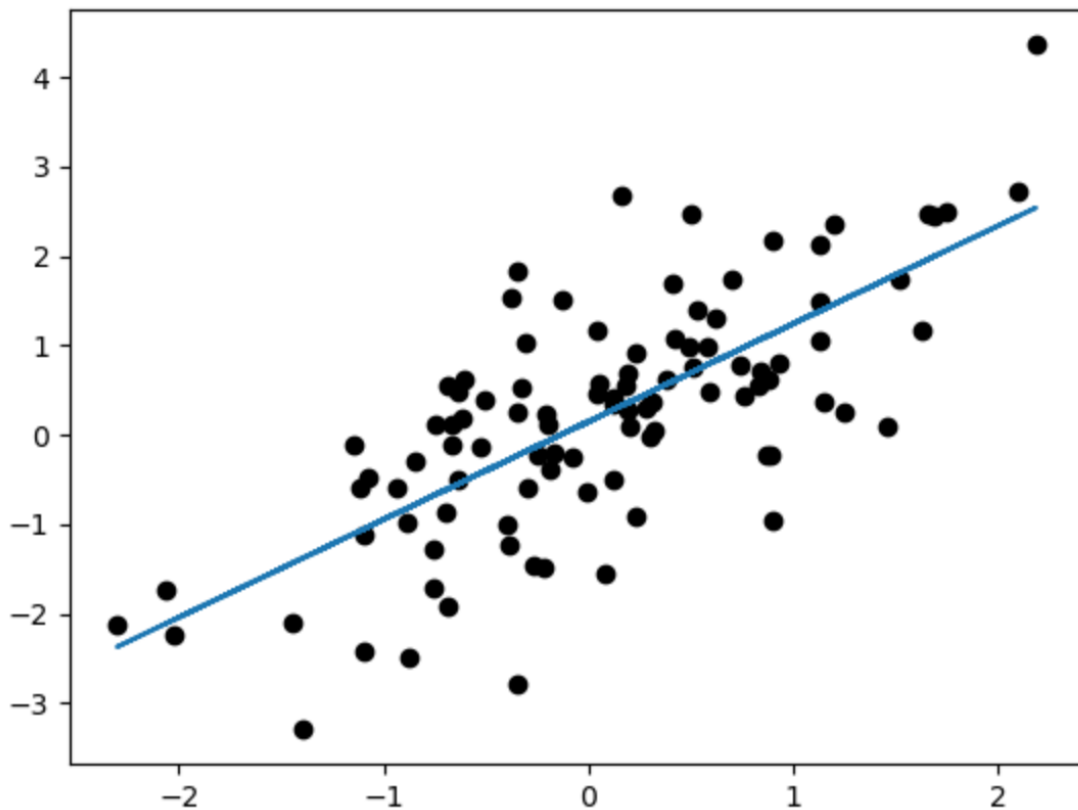
回帰

回帰とは

回帰とは機械学習のうちの「教師あり学習」の1つで、「売り上げ」や「成長率」といった数量を扱う場合の学習方法です。例えば、過去の顧客データから新しい客が今後何回訪れるのかを予測するといったことです。

線形回帰

線形回帰は非常にシンプルな手法で、データ分布に最も当てはまる直線を考えるものです。イメージとしては以下の画像のようになります。以下の画像の直線を元に、例えば横軸が「3」の時の縦軸の数値を予測することができます。



Orizuru 記事より引用

サポートベクターマシン (SVM)

分類のパートで既に紹介したのであまり深くは説明しないですが、**分類**と**回帰**では計算方法は大きく異なるため、同じ**SVM**といっても異なる部分が多いです。

SVM分析に本当に必要なデータ（サポートベクトルという）のみを上手に利用するという発想の下、回帰問題における**SVM**では予測結果の誤差をどう解釈するのかといった工夫をすることもあります。

サポートベクター回帰 (SVR)

サポートベクター回帰はサポートベクターマシンの回帰版で、多くの変数を含んだ非線形な問題を解決することができる手法です。

SVMがマージンの最大化と誤推計サンプルを減らすやり方で学習するように、**SVR**では重みの最小化と誤差の最小化で学習します。

誤差の小さい部分を無視することで、分析結果のロバスト性を向上させています。

Elastic Net

Elastic Netは**ラッソ回帰**と**リッジ回帰**という2つの手法を合成した手法です。

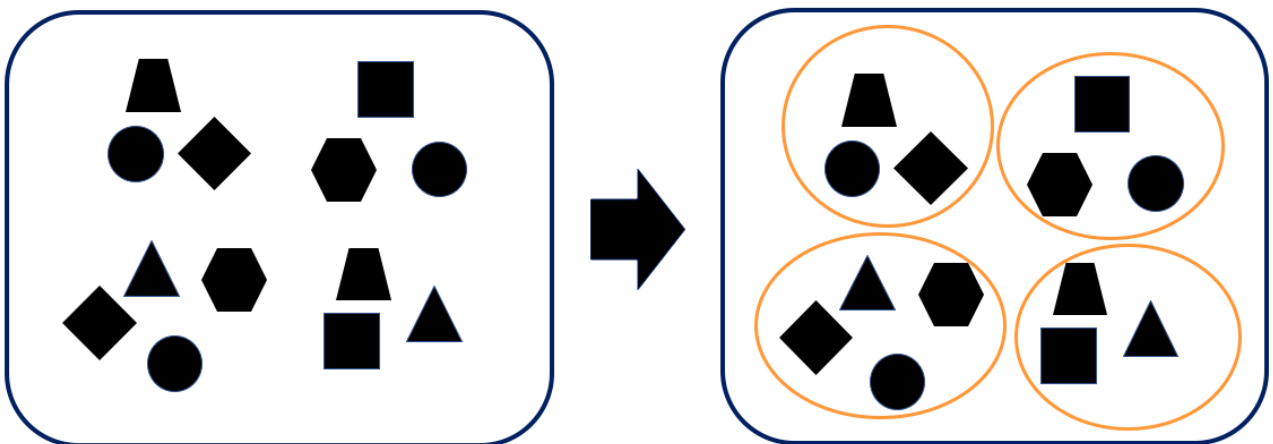
ラッソ回帰もリッジ回帰も機械学習でしばしば起きる**過学習**を防ぐために行う**正則化**という方法を適用した機械学習で、それら2つを合わせた**Elastic Net**はロバスト性も、予測性能も高い手法になっています。

クラスタリング

「クラスタリング」とは**分類**の延長上にある手法で、「似たデータの集まりを機能やカテゴリごとに分けて集める」代表的な教師なし学習手法です。

イメージとしては以下の画像のようなものです。

クラスタリングのイメージ



AINOW編集部作成

データの全体集合を**クラスタ**という部分集合に分けることであり、内的結合と外的分離の性質を持ちます。クラスタリングでは基本的に何かしらの指標で近い距離にあるもの同士を1つのクラスタと見なすことが多いです。

クラスタリングのアルゴリズムには**階層クラスター分析**と**非階層クラスター分析**の2種類があります。

階層クラスター分析

階層クラスター分析は階層構造のある分類方法です。最も似ているデータの組み合わせからクラスター（ひとまとまり・グループ）に分けていく方法です。

また出力されたクラスターを樹形図で見することもできますが、大量のデータを分類することには向いていないため、対象の少ないデータに実行するのに有効です。

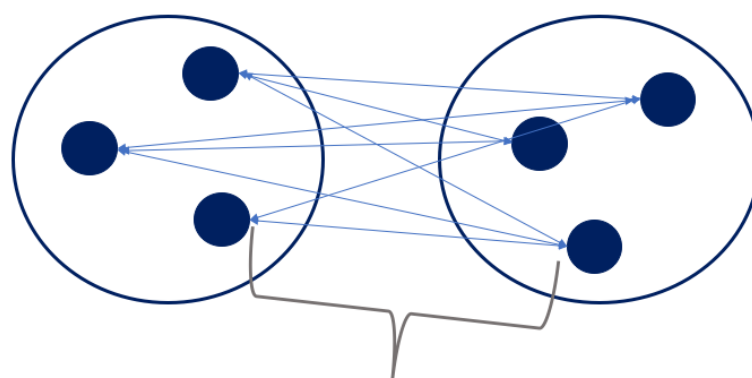
ここでは群平均とWard法について説明します。

Ward法

Ward法では既にあるクラスターのうち、一番近い2つのクラスターを合わせることでクラスターリングします。データ全体がバランスよく分類されやすいので、比較的によく利用されています。

群平均

2つのクラスター内のデータ同士の距離を比較して、クラスター間の距離を定める手法です。この方法を利用することで**鎖効果**という余分なクラスターが生まれてしまう現象を防止できるのです。以下の画像がイメージ図です。



この距離が各クラスターを定義づける

AINOW編集部作成

非階層クラスター分析

非階層クラスター分析は階層構造のない分類方法です。

事前にクラスター数を決めておき、その数だけのクラスターにデータを分類します。階層クラスター分析に比べて、ビッグデータなどの大量のデータに実行するのに適しています。ここでは**k-means法**について説明します。

k-means

k-means法とは、「クラスターの平均（means）」を用い、あらかじめ決められたクラスター数“k”個に分類する」ことに由来しています。

k個のベクトルを利用することから始まり、ベクトルを調整しながらデータを分類していきます。ちなみにデータ分析のソフトウェア大手である**Tableau**の分析の一部にも**k-means法**は利用されています。

次元削減

機械学習でも特徴量が不必要に多すぎると、いわゆる「次元の呪い」という現象が起こり、精度が悪くなることがあります。次元削減はデータの次元（特徴量の数）を減らす手法です。次元削減を行う目的は主に以下の二つで、ここでは主成分分析（**PCA**）について紹介します。

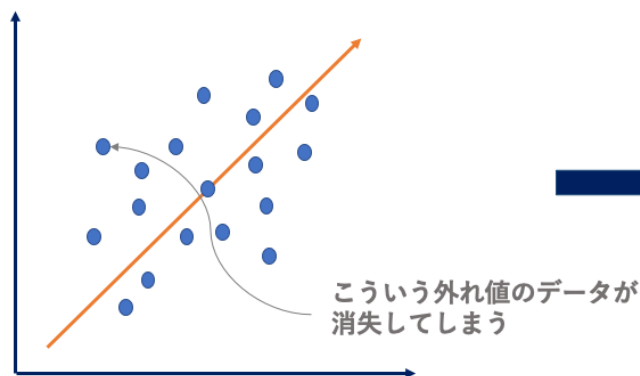
データの圧縮

データの可視化

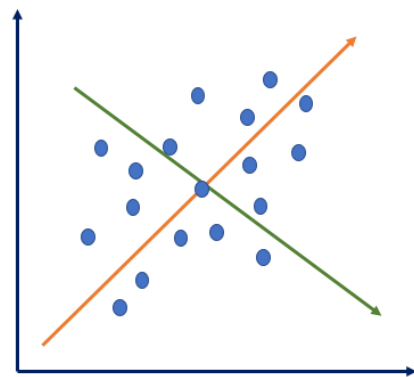
主成分分析（PCA）

主成分分析とは、複数の変量データを統合し、新たに1つの総合指標を作るための手法です。つまり、クラスター分析のように多数のデータを「グループ分け」して全体の見通しを良くする方法です。

軸を設定し、データ削除



新しい軸で情報の損失を補う



AINOW編集部作成

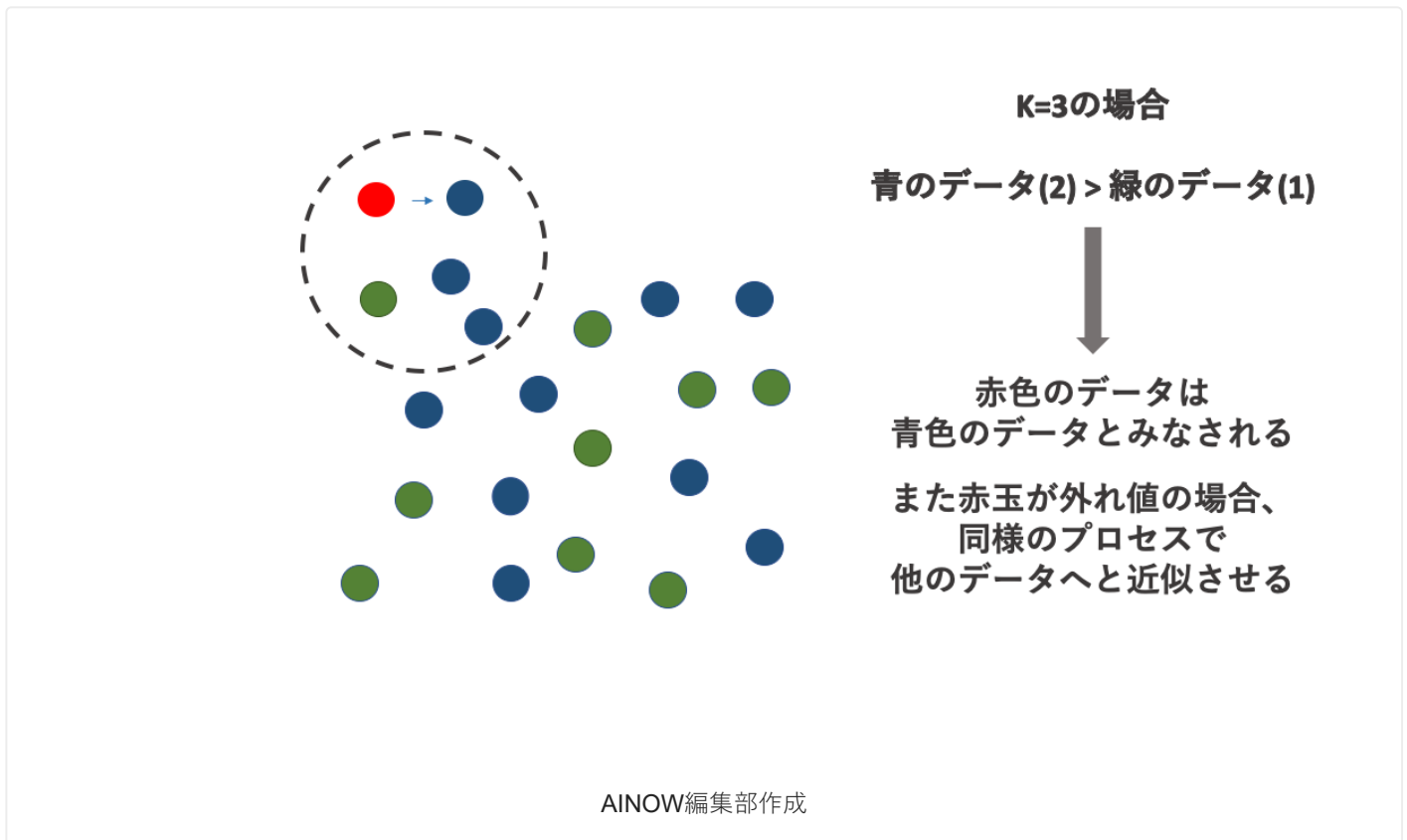
異常検知

異常検知とは機械の故障やデータ分析の外れ値を検知・推測する際に利用する手法です。

異常検知は、データセット内の他のデータと一致していない観測結果や期待されるパターン等をデータマイニングなどで識別することを言います。また外れ値検知という異常検知では正常時には生まれない外れ値を見つけ出すために機械学習が使われたりします。ここでは3つ紹介します。

k近傍法

k近傍法は新しいデータや判別したいデータに対して、あらかじめ決められた**k個**のデータが作る範囲内と多数決の原理を利用して分類のようなことをする手法です。



主成分分析

主成分分析は次元削減のパートで紹介されているので説明は省略しますが、主成分分析を利用することで、異常値（外れ値）をデータの値が完全に消失しない程度に他の軸に代表してもらい、他のデータの値に近似させる方法で外れ値を実質的に無くします。

One-class SVM

One-class SVMは文字通り、SVMの手法を応用して識別境界を設定して、通常値と外れ値を識別する**外れ値検知**のための手法と言えます。

異常がなかなか発生せず、異常データが集まらないような状況でも異常検知を行いたい場合には有効な手法と言えます。