

機械学習の概要

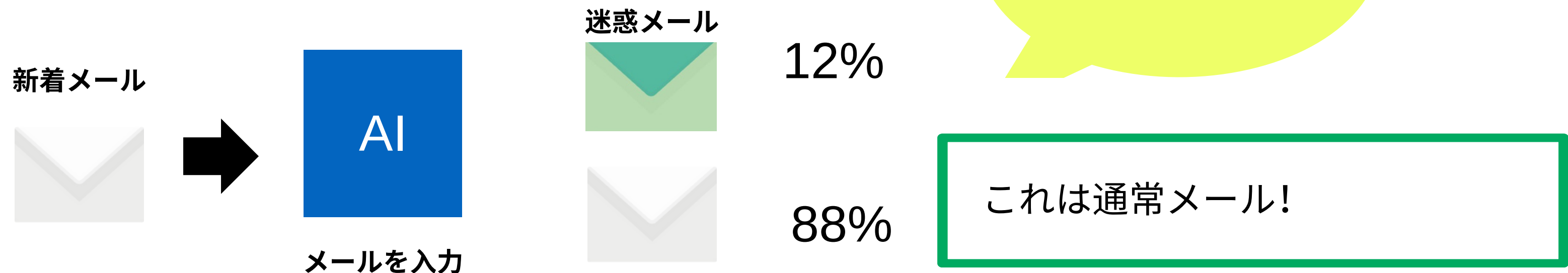
本講座のゴール

- 本講座は8部構成のうち2回目です。
- Part2は「機械学習の概要」です。教師あり、なし学習、強化学習、各種機械学習の手法、モデルの評価について理解いただきます。
- G検定を受験しようとしている方向けに、頻出項目に絞って解説し、章末問題を解きながら合格に必要な知識を身につけていただきます。

機械学習とは

- ・ 機械学習（machine learning）とは、人間が自然に行なっている学習能力と同様の機能をコンピュータで実現しようとする技術や手法。
- ・ **アーサー・サミュエルは、明示的にプログラムしなくても学習する能力をコンピュータに与える研究分野と定義。**
- ・ 機械学習を活用することで、分類（迷惑メールか通常メールか）や、回帰（株価予測や売上予測）などができる。

例えば、機械学習の分類の場合、大量の通常メールと迷惑メールを学習させ、学習後新着メールが入力されると、通常メールか迷惑メールかを予測する。



機械学習の種類

機械学習は大きく3つに分けられる。

○ 教師あり学習

問題と正解のペアを与えて学習する方法

○ 教師なし学習

学習データに正解をつけずに学習する方法

○ 強化学習

試行錯誤し「価値を最大化するような行動」が何かを学習する方法。とるべき行動を最適化します。

教師あり学習と教師なし学習を組み合わせた「半教師あり学習」もあります。
4種類について解説します。

教師あり学習

教師あり学習は『回帰』と『分類』に分けられます。

部屋の広さ	駅からの距離	築年数	家賃
1LDK	5	3	63000
2LDK	20	8	58000
1LDK	3	3	60000
2LDK	30	30	55000

説明変数
(目的変数を予測・説明するための手がかりとなる変数)

回帰問題
連続値を予測
住宅価格予測など

目的変数 (予測したい値)

年齢	性別	職業	融資承認
52	1	無職	0
25	1	エンジニア	1
28	0	営業	1
34	0	採用担当	1

分類 (識別) 問題
カテゴリを分ける
融資が承認されるか、されないか。
犬や猫の分類や
メールのスパムフィルタリング

教師なし学習

教師なしは目的変数（正解）を用いない機械学習手法です。
教師なし学習には、**次元削減**や**クラスタリング**などがあります。

- ・次元削減とは、データの表現を変えることによって、データの情報をできるだけ失うことなくデータを低い次元に圧縮する。
 - ・代表的な手法の1つとして**主成分分析**がある。
 - ・クラスタリングとは、データをいくつかのクラスタ（グループ）に分ける。
 - ・代表的な手法の1つとしてK-means法がある。
-

強化学習

強化学習

- ・ 試行錯誤して「価値を最大化するような行動」が何かを学習する。
- ・ 教師あり学習とよく似た問題設定だが、与えられた正解の出力をそのまま学習すれば良いわけではなく、一言で言えば、「価値」を最大化する行動を学習する。
- ・ 「価値」を最大化することが強化学習の目的。
- ・ 囲碁や将棋といったゲームは、本質的に将来の価値（つまり今その手を打つことにより、最終的に勝つのか負けるのか）を最大化することが目的なので、強化学習とは相性が良い（プロの囲碁棋士を圧倒したAlphaGoなど）。

Q学習

- ・ 強化学習のアルゴリズムの1つ。
 - ・ 報酬を最大化させる方策を見つける
-

半教師あり学習

- ・ 半教師あり学習とは、教師あり学習と教師なし学習を組み合わせた手法。
 - ・ 学習データに、正解ラベルのついたデータとついていないデータの両方を用いる。
 - ・ 最初是一部のデータにのみ正解ラベルが付与されていて、それらを用いて残りのデータに対してラベルを事前に予測する。
最後に全てのデータをあわせる。
 - ・ 正解ラベルのデータを十分に用意できない場合に利用される。
-