

Python ではじめる機械学習

1. はじめに

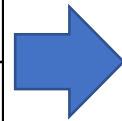
1.1 なぜ機械学習なのか？

15T4057F 藤井 真

1. はじめに

- 機械学習とは
 - データから知識を引き出すこと

工学部.data		
No.	男1女0	眼鏡
1	1	1
2	0	0
3	1	0
4	1	1
5	1	1
6	1	1
7	1	1



工学部には眼鏡の男が多い

1. はじめに

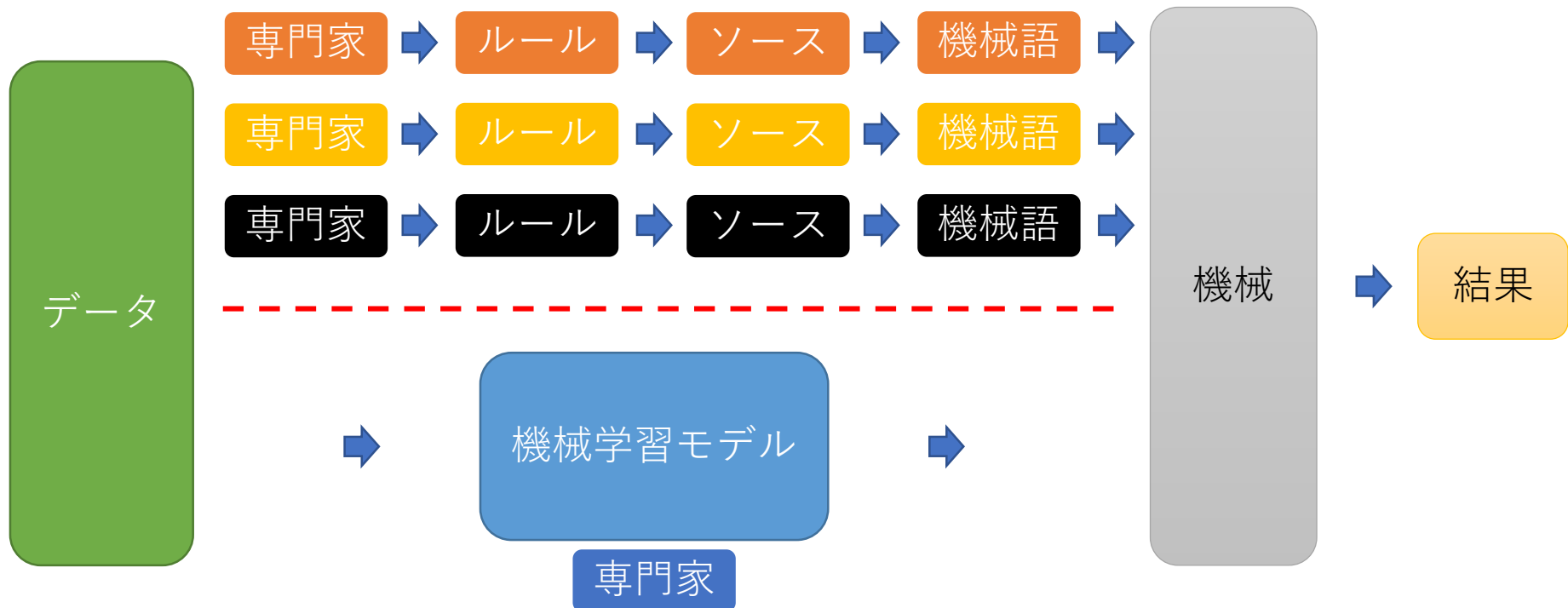
- 機械学習とは
 - データから知識を引き出すこと
 - 統計学、人工知能、計算機科学が交差する領域
 - 予測解析、統計学習とも呼ばれる
 - 例を挙げると
 - レコメンド（おすすすめ、推薦）機能
 - 画像認識
 - 惑星観測
 - 素粒子の発見
 - DNA解析
- etc.

1.1 なぜ機械学習なのか？

- これまでの「知的」アプリケーション
 - 「知的」の内容について詳しい専門家が関わり、「もし～なら」という形式で記述する。
- 問題点
 - 専門性：「知的」の内容をよく理解した専門家が不可欠。
 - 汎用性：「知的」の内容が変われば書き直す必要がある。
 - 人と計算機の知覚は内部表現レベルで違う。
 - 人がルールを機械に教え続けるには限界がある。
 - 機械自身が「知的」について学習した方が良いのでは？

• 問題点

- 専門性：「知的」の内容をよく理解した専門家が不可欠。
- 汎用性：「知的」の内容が変われば書き直す必要がある。
- 人と計算機の知覚は内部表現レベルで違う。
- 人がルールを機械に教え続けるには限界がある。
- 機械自身が「知的」について学習した方が効率の良い分野があるのでは？



1.1.1 機械学習で 解決可能な問題

- 問題「男女の判別」

- データ「family_name」：鈴木、佐藤、張、劉、etc.
⇒判別のための特徴や情報がない⇒解決不可能
- データ「first_name」：花子、太郎
⇒統計的に男女に偏りがある、判別できる⇒解決可能

つまり、

データに情報（特徴や傾向）が無ければ解決できない。

教師あり学習と教師なし学習

- 教師あり学習

①人が望ましいとする入出力のペア群を機械に与える。（望ましい状態を教える＝教師）

②機械はその二つから学習し、モデルを作成し、未知の入力に対しても人の助けなしに出力を生成する。

デメリット：人が答えを教え続けなければならない

⇒コストがかかる

例：封筒の郵便番号、顔認識、不正検出

教師あり学習と教師なし学習

- 教師なし学習

- ①機械には入力データだけが与えられ、人が望ましいとする出力などは与えられない。
- ②機械はデータから学習し、モデルを作成し、未知の入力に対しても人の助けなしに出力を生成する。

ただし、望ましい状態（教師）を与えていないので出力の表現形式や内容が人の望む結果から遠い場合がある。

例：テキストのトピック抽出、顧客嗜好
webへの異常アクセスの検出

機械へのデータの渡し方 (一般的なデータ形式)

<列>
・特徴量

<行>
・サンプル
・データポイント

No.	男1女0	眼鏡	学科	学年	単位
1	1	1	4	1	22
2	0	0	4	2	32
3	1	0	3	3	60
4	1	1	2	4	50
5	1	1	5	5	80
6	1	1	1	6	102
7	1	1	1	7	111

1.1.2 タスクを知り、 データを知る

- 機械学習において最重要

扱うデータを理解すること

解決したい「問題」と「データ」の関係を理解

- 答えようとしている問題は何なのか
- 機械学習が適しているか、
- 集めたデータで答えられるのか、質、量は十分か
- うまく答えられたかをどのように判断するのか

機械学習（手段）がしたいのではなく、問題が解きたい（目的）。
機械学習は問題を解くための過程の一部であり、一手段。
常に問題全体を意識する必要がある。