

# 情報リテラシー（第11回 後期）

## 機械学習の基礎－3つの学習方法－

# 今日のねらい

- 機械学習の全体像を理解できる
- 教師あり学習、教師なし学習、強化学習の違いを説明できる
- それぞれの具体例を知る
- どの手法をどんな問題に使うか判断できる

# Week 10の振り返り

## 前回学んだこと

- 仮説検証サイクル
- 線形回帰（単回帰）
- 打率 → 本塁打を予測
- train/test分割（60:40）

## 今日は

機械学習の全体像を学びます  
線形回帰も機械学習の一種です

# 機械学習とは 定義

データから学習し、予測や判断を行う技術  
従来のプログラミングとの違い

**従来：**ルールを人間が書く  
例：「打率が.300以上なら強打者」

**機械学習：**データからルールを学習  
例：データを見て「強打者の特徴」を発見

# 機械学習の3つの分類

## 1. 教師あり学習 (Supervised Learning)

正解ラベル付きデータで学習

## 2. 教師なし学習 (Unsupervised Learning)

正解ラベルなしでパターンを発見

## 3. 強化学習 (Reinforcement Learning)

試行錯誤で最適な行動を学習

教師あり学習とは  
データの形

説明変数 + 正解ラベル

例：

- 画像 + 「猫」「犬」のラベル
- メール + 「スパム」「正常」のラベル
- 前回：打率 + **本塁打の実測値**

目的

新しいデータに対して**正解を予測**

# 教師あり学習：回帰と分類

## 回帰 (Regression)

### 数値を予測

例：

- 打率 → 本塁打数 (Week 10)
- 気温 → アイスクリームの売上
- 広告費 → 売上

## 分類 (Classification)

### カテゴリを予測

例：

- メール → スパム or 正常
- 画像 → 猫 or 犬
- 打率 → 本塁打30本以上 or 未満

# 教師あり学習の具体例

## スパムメール判定

メールの内容から「スパム」か「正常」かを判定

## 画像認識

写真から「猫」「犬」「鳥」などを識別

## 株価予測

過去のデータから将来の株価を予測

## 採用判定 (Week 9)

履歴書から「採用」「不採用」を判定

# デシジョンツリーとは 決定木 (Decision Tree)

## 質問を繰り返して分類

例：本塁打30本以上かどうか

```
打点が70以上?  
YES → 30本以上  
NO  → 打率が.300以上?  
YES → 30本以上  
NO  → 30本未満
```

## メリット

- 視覚的に分かりやすい
- 判断理由が明確

# 教師なし学習とは データの形

説明変数のみ（正解ラベルなし）

例：

- 顧客の購買データ
- 選手の成績データ
- Week 10と同じデータだが、ラベルを使わない

目的

データの中からパターンを発見

# 教師なし学習：クラスタリング

## クラスタリング (Clustering)

似ているものをグループ化

例：

- 顧客をセグメント分け
- 選手を3つのグループに分類
  - 強打者グループ
  - 中堅グループ
  - 弱打者グループ

**重要：**正解ラベルなしで自動的にグループ化

## 教師なし学習の具体例

### 顧客セグメンテーション

購買パターンから顧客を分類  
→ マーケティング戦略に活用

### 異常検知

通常と異なるパターンを見つける  
→ 不正検知、機器の故障予測

### レコメンドシステム

似た商品をグループ化  
→ 「この商品を買った人はこれも買っています」

# K-meansとは

## K-means法

最も基本的なクラスタリング手法

### 仕組み

1. K個のグループを作る (Kは事前に決める)
2. 各データを最も近いグループに割り当てる
3. グループの中心を再計算
4. 繰り返し

### 今日の実習

**K=3** (3グループ) で選手を分類

# 強化学習とは 定義

試行錯誤で最適な行動を学習

## 仕組み

1. 行動：何かをやってみる
2. 観察：結果を見る
3. 報酬：良ければ+、悪ければ-
4. 学習：次はもっと良い行動を選ぶ
5. 繰り返し

## イメージ

ゲームで上手くなる過程

## 強化学習の具体例

### AlphaGo（囲碁AI）

- 2016年に人間のプロ棋士に勝利
- 何千万回もプレイして学習
- 勝ったら報酬、負けたら罰

### ロボット制御

- 二足歩行ロボットが転ばずに歩く
- 試行錯誤でバランスを学習

### 自動運転

- 安全に目的地まで運転
- シミュレーションで何度も練習

# 3つの学習方法の比較

|     | 教師あり    | 教師なし    | 強化学習    |
|-----|---------|---------|---------|
| データ | 正解あり    | 正解なし    | 報酬あり    |
| 目的  | 予測      | パターン発見  | 最適な行動   |
| 例   | スパム判定   | 顧客分類    | ゲームAI   |
| 野球例 | 30本以上予測 | 選手グループ化 | 打撃戦略の学習 |

どれを使う？

- **予測したい** → 教師あり
- **パターン発見** → 教師なし
- **行動を最適化** → 強化学習

# 今日のColab実習

## Part 1：教師あり学習（デシジョンツリー）

打率 + 打点 → 本塁打30本以上を予測

## Part 2：教師なし学習①（成功例）

打率 + 打点でクラスタリング（3グループ）

## Part 3：教師なし学習②（失敗例）

打率 + 出塁率でクラスタリング（3グループ）

## Part 4：比較

なぜ結果が違うのか？

# 変数選択の重要性

## 成功例：打率 + 打点

→ グループ1が**100%完璧**に30本以上を発見

## 失敗例：打率 + 出塁率

→ **90%程度**で不完全

## なぜ？

打率と出塁率は**似た変数**（相関0.875）

→ 新しい情報が少ない

## 教訓

- 目的変数と**相関が強い変数**を選ぶ
- **異なる情報**を持つ変数を組み合わせる

# 課題

今回は課題なし！

授業内で実習を完結させます

# まとめ

## 機械学習の3つの学習方法

- 教師あり : 正解ラベルで学習 → 予測
- 教師なし : パターン発見 → グループ化
- 強化学習 : 試行錯誤 → 最適な行動

## 今日の体験

- デシジョンツリーで分類
- K-meansでクラスタリング
- 変数選択の重要性

## 振り返り :

Teamsに今日の感想を書いてください