



## 教師なし学習

教師なし学習（**unsupervised learning**）は、ラベルなしデータを扱う機械学習である

ref: なっとく！機械学習 p25～26

ラベル（予測の目的変数または正解値）がないデータから、できるだけ多くの情報を抽出することが目標となる

たとえば、ラベルが付いていない動物の画像のデータセットからは、それぞれの画像が表している動物の種類はわからないため、新しい画像がどの動物なのかを予測することはできない

しかし、2つの画像が似ているかどうかなど、他にできることがある

つまり、教師なし学習アルゴリズムは、類似性に基づいてデータを分類できるが、それぞれのグループが何を表すのかはわからない



## 教師なし学習によるデータの前処理

実際には、教師なし学習はラベルが付いている場合でも利用できる

ref: なっとく！機械学習 p26

教師なし学習を使ってデータの**前処理**を行うと、教師あり学習の手法の効果を高めることができる



## 教師なし学習の種類

教師なし学習には、大きく分けて 3 種類の学習法がある

ref: なっとく！機械学習 p26

- **クラスタリング**：データを類似性に基づいてクラスタに分類する
- **次元削減**：データを単純化し、より少ない特徴量でデータを正確に説明する

- **生成学習** : 既存のデータに似ている新しいデータ点を生成する



## クラスタリング

**クラスタリング** (**clustering**) は、データセット内の要素を類似性の高いデータ点ごとにクラスタ (グループ) に分割する

ref: なっとく! 機械学習 p26~30

特徴量が 3 つを超えると、その次元を可視化できなくなるため、人間がクラスタを目で確認するのは不可能になる

コンピュータを使うことで、巨大なデータセットに対してもクラスタリングを行うことができる



## 次元削減

[ Todo 1: ]

ref: なっとく! 機械学習 p30~32



## 行列分解と特異値分解

[ Todo 2: ]

ref: なっとく! 機械学習 p32~34



## 生成学習

[ Todo 3: ]

ref: なっとく! 機械学習 p34

.....



# Zebra Notes

Type	Number
todo	3