Seminar
ベイズ推論による機械学習　入門

2024/05/22

Nango Lab.

M2 Fukuda Akiya

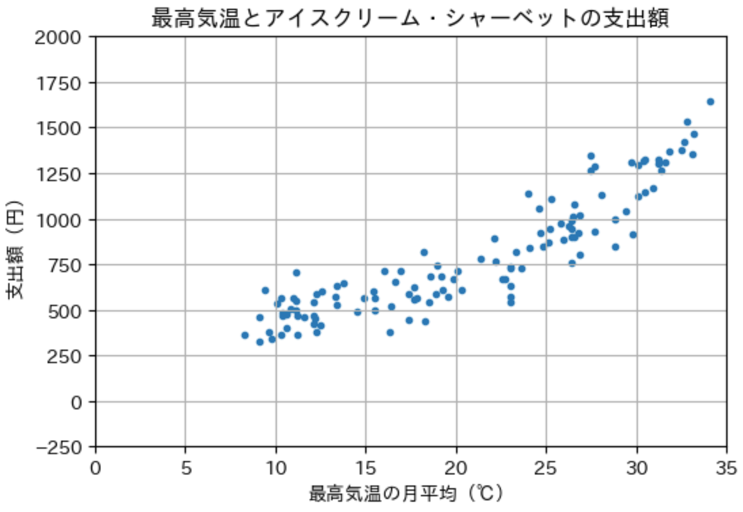
機械学習とは？

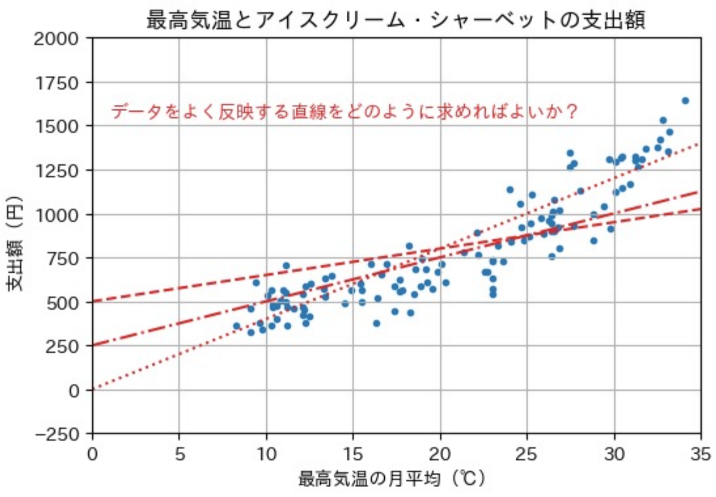
Wikipedia

人工知能における研究課題の一つで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・手法

本書

データに潜む規則や構造を抽出することにより、
未知の現象に対する予測やそれに基づく判断を行うための計算技術の総称





機械学習帳 https://chokkan.github.io/mlnote/index.html

機会学習の代表的なタスク

- **回帰（Regression）**
  離散値$x \in \mathbb{R}^N$から連続値$y = f(x)$を求める

- **分類（Classification）**
  出力値yを有限個のシンボルに限定するモデル
  例 : 画像に猫が写っているか否か

- **クラスタリング（Clustering）**
  与えられたN個のデータを何かしらの基準にしたがって、K個の集合に分ける
  ポワソン混合モデル、ガウス混合モデル（4章）

- **行列分解（Dimensionality Reduction）**
  ある行列で表されるデータを二つの行列の内積で表す

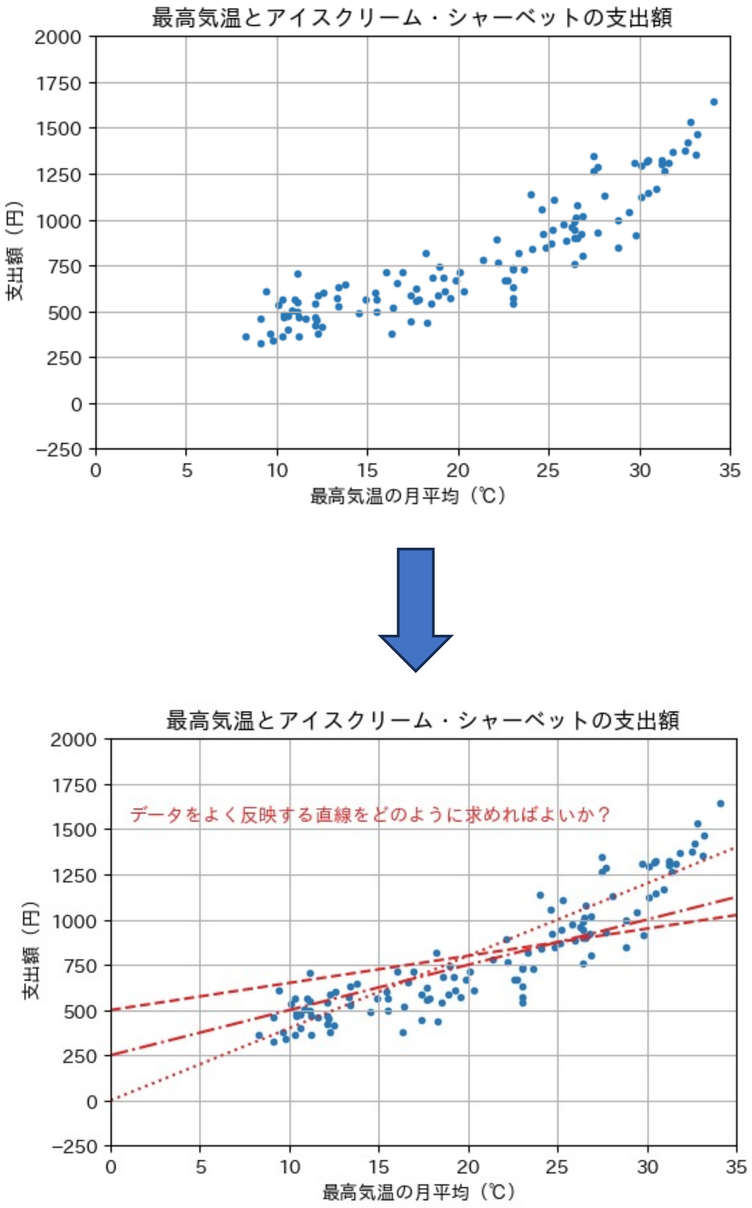
回帰

- 回帰（Regression）
  離散値$\boldsymbol{x} \in \mathbb{R}^N$から連続値$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$を求める

- 入力データ：$\boldsymbol{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$
  出力データ：$\boldsymbol{Y} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\}$
  重み付きパラメータ：$\boldsymbol{w} \in \mathbb{R}^M$
  ノイズ：$\boldsymbol{\epsilon}_n$
  線形回帰（linear regression）モデル

$$\boldsymbol{y}_n = \boldsymbol{\omega}^T \boldsymbol{x}_n + \boldsymbol{\epsilon}_n \quad (1.1)$$

- 各変数$\boldsymbol{x}, \boldsymbol{y}$の関係性を再現するような$\boldsymbol{\omega}$を求める



4

機械学習帳 https://chokkan.github.io/mlnote/index.html

分類

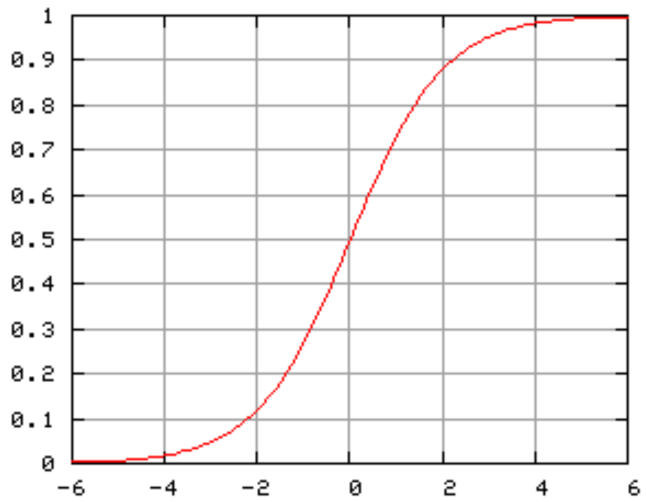
- 分類（Classification）
  出力値yを有限個のシンボルに限定するモデル

- 出力値 : $y_n \in \{0, 1\}$とし、y_nが1ととる確率を$\mu_n \in \{0, 1\}$
  M次元の入力値x_n, パラメータωから、μ_nを表現するモデル

$$\mu_n = f\left(w^T x_n\right) \quad (1.3)$$

- 関数fにはシグモイド関数（Sigmoid function）がよく用いられる

$$f(a) = \mathrm{Sig}(a) = \frac{1}{1 + e^{-a}} \quad (1.4)$$



Wikipedia, シグモイド, https://ja.wikipedia.org/wiki/シグモイド

機械学習の2つのアプローチ

- ツールボックスとしての機械学習
  既存の様々な予測アルゴリズムに対してデータを与え、その中から何らかの基準に従って性能の良いアルゴリズムを選ぶことによって最終的な予測や判断
  ▶高度な数学的知識がなくても、多少のプログラミング知識で利用可能

  例：最近傍法、サポートベクターマシン、ブースティングなど…
  ▶入力データと出力データ（正解ラベル）のペアを訓練データとして予測のための関数を学習するので, 教師あり学習と呼ぶ

- モデリングとしての機械学習
  データに関するモデル（仮説）を事前に構築し, モデルの含むパラメータや構造をデータから学習することによって, 予測や判断

  例：時系列モデル、状態空間モデル
  ▶データの離散的な変化をモデル化するための隠れマルコフモデル
  ▶状態の連続的な変化をモデル化するための線形動的システム

確率の基本計算

- **確率分布**
  各要素が連続値であるような$\boldsymbol{M}$次元ベクトル$\boldsymbol{x} = (x_1, \dots, x_M)^T \in \mathbb{R}^M$に対する関数$p(\boldsymbol{x})$が次の2つを満たすとき, $p(\boldsymbol{x})$を確率密度関数という

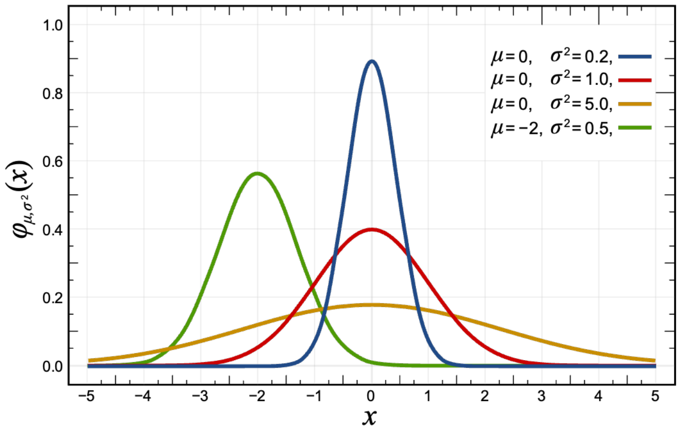
$$p(\boldsymbol{x}) \geq 0 \quad (1.8)$$

$$\int p(\boldsymbol{x}) d\boldsymbol{x} = \int \dots \int p(x_1, \dots, x_M) dx_1 \dots dx_M = 1 \quad (1.9)$$

- **例：$\boldsymbol{M} = 1$次元のガウス分布に対する確率密度分布**

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (1.10)$$

平均：$\mu \in \mathbb{R}$、**分散：**$\sigma \in \mathbb{R}$



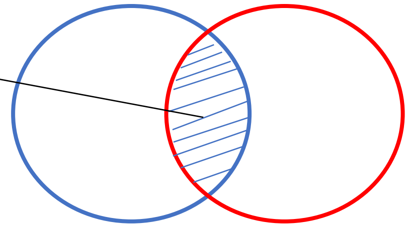
- **各要素が離散値の場合**

$$\sum_x p(x) = \sum_{x_1} \dots \sum_{x_M} p(x_1, \dots, x_M) = 1 \quad (1.12)$$

Wikipedia, 正規分布, https://ja.wikipedia.org/wiki/正規分布

確率分布の基本式

- 同時分布$p(x,y)$：ある2つの変数x,yに対する確率分布



- 周辺分布：

$$p(y) = \int p(x,y)dx \quad (1.16)$$

一方の変数を除去する操作を周辺化という

- 条件付き分布：yに対して特定の値が決められた時のxの確率分布

$$p(x|y) = \frac{p(x,y)}{p(y)} \quad (1.17)$$

式(1.16)、(1.17)より

$$\int p(x|y)\,dx = \frac{\int p\,(x,y)dx}{p(y)} = 1 \quad (1.18)$$

ベイズの定理

- xが与えられたときのyの条件付き確率

$$p(y|x) = \frac{p(x,y)}{p(x)} \quad (1.19)$$

式(1.17)より

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p\,(x,y)dx} \quad (1.20)$$

ベイズの定理

原因xから結果yが得られる確率$p(y|x)$から、結果yが得られたときの原因xの確率$p(x|y)$を逆算するような手続き

- 同時分布の独立性

$$p(x,y) = p(x)p(y)(1.21)$$

両辺を$p(y)$で割ると

$$p(x|y) = p(x)(1.22)$$