

# 1. 正規化 (Normalization)

学習に使用するパラメータを  $0 \leq x \leq 1$  の範囲に収まるようにする処理のこと.

具体例 :

例えば, 物価予測をする際に土地の面積や部屋数, 階数, 駅からの距離などの単位が異なる特徴量をごちゃまぜで学習させると, 収束が遅く, なかなか最小値に近づかない. これは, 土地の面積が 0 200平米, 部屋数が0 5部屋など**値ごとにばらつきが大きい**から. そこで, それぞれの特徴量を  $0 \leq x \leq 1$ の範囲になるようにする.

## 1.1. Mean Normalization

$$x := \frac{x - \text{average}}{\text{max} - \text{min}}$$

## 1.2. min-max noemalization

$$x := \frac{x - \text{min}}{\text{max} - \text{min}}$$

## 1.3. 平均を 0, 分散を 1 にする正規化

確率変数 (またはデータ)  $X$  に対して,

$$Y = \frac{X - \mu}{\sigma}$$

と変換すると,  $Y$ の平均は 0, 分散 (および標準偏差) は 1 となる.

## 2. 正則化 (Regularization)

正則化という言葉の由来は、線形代数の正則行列 (Regular Matrix) らしい。

機械学習における正則化とは、過学習 (Over-fitting) の回避が目的。

回帰モデルのコスト関数  $J(\theta)$  は以下のようになる (二乗和誤差)。

$$J = \frac{1}{2m} \sum_{i=1}^m \|h_{\theta}(x^i) - y^i\|^2$$

この式に正則化を加えると、

$$J = \frac{1}{2m} \sum_{i=1}^m \|h_{\theta}(x^i) - y^i\|^2 + \lambda \sum_{j=1}^n (\theta_j)^2$$

というようにパラメータ  $\theta$  の二乗和を追加する。この項を正則化項と呼ぶ。これにより、 $\theta$  の値がどれかに偏ったりすることを防ぎ、満遍なく  $\theta$  に値があるように抑制させておくことができる。これが過学習を防ぐ働きになる。

## 2.1. 正則化項

### 2.1.1. Lasso Regularization

$$\sum_{j=1}^n |\theta_j|$$

Lasso は, Least absolute shrinkage and selection operator の略で **L1 正則化** という.

### Ridge Regularization

$$\sum_{j=1}^n \theta_j^2$$

この正則化は **L2 正則化** という. さらに, L0 正則化は  $\max(\theta_j)$  となるほか, Lasso と Ridge の合わせ技などもある.