

Three bright green apples are arranged in a cluster on a white background. One apple is in the foreground, slightly to the right, and two are behind it, one to the left and one to the right. The apples are glossy and have short brown stems.

人工知能

IE229 – ARTIFICIAL INTELLIGENCE

第3回講義: 線形回帰

Lecture 3: Linear Regression

二宮 崇 (Takashi Ninomiya)

愛媛大学 (Ehime University)

ninomiya@cs.ehime-u.ac.jp

【復習】機械学習

- データから関数を学習

- 関数は、入力(x)と出力(y)の関係を表す

$$x \xrightarrow{f} y$$

- 大量の入出力ペア(x, y)の集まり(データ)から関数 f を予測

大量のデータ

(x_1, y_1)

(x_2, y_2)

(x_3, y_3)

\vdots

(x_n, y_n)

学習

f



【復習】機械学習

● データから学習

- 入力 $\mathbf{x} = (x_1, x_2, \dots, x_m)$ ← m 次元ベクトル
(画像の場合は、赤、青、緑の画像に対応する3つの行列(テンソル))
- データ D は入力 \mathbf{x} と出力 y のペアの集合
- 関数 f は **パラメータ(重み変数)の集合** $\mathbf{w} = (w_1, w_2, \dots, w_m)$ から成る
例: $f(\mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_mx_m$
 - f は $f_{\mathbf{w}}(\mathbf{x})$ と書くとわかりやすい
- 学習 = データ D に対する誤差(損失)を最小にするパラメータ \mathbf{w} を求める

データ全体の誤差(損失)
$$L(\mathbf{w}) = \sum_{(\mathbf{x}, y) \in D} (y - f_{\mathbf{w}}(\mathbf{x}))^2$$



【復習】 回帰問題と分類問題と構造予測

関数: $y = f(x)$

- 回帰問題 (regression)

- $y \in R$ (実数)

例: 年齢予測、降水確率の予測、気温の予測

- 分類問題 (classification)

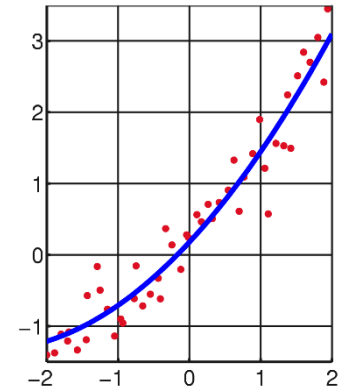
- $y \in \{C_1, C_2, \dots, C_K\}$ (ラベル集合)

例: 文書分類(政治、経済、スポーツ等)

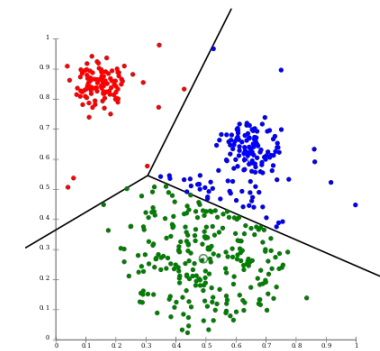
- 構造予測 (structured prediction)

- $y \in G$ (グラフ集合)

回帰

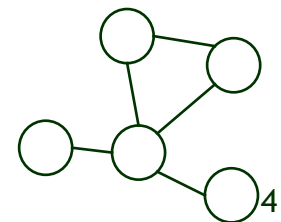
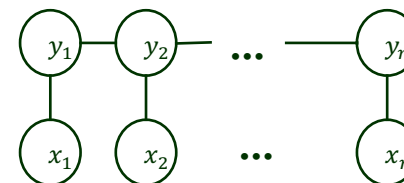


分類



© Chire CC BY-SA 3.0

構造予測



損失関数

- 損失関数 $L(y, \hat{y})$
 - 誤差を数値化する関数。 y は正解、 \hat{y} は予測値

- データ全体の損失 $L(\mathbf{w})$

$$L(\mathbf{w}) = \sum_{(x,y) \in D} L(y, f(\mathbf{x}))$$

学習：データ全体の損失が最小になる各パラメータの値を探すこと

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} L(\mathbf{w})$$

- 回帰問題の損失関数(二乗損失)

$$L(y, \hat{y}) = (y - \hat{y})^2$$

- 分類問題の損失関数(負対数尤度、交差エントロピー損失)

$$L(y, \hat{y}) = -\log p_{\mathbf{w}}(y|\mathbf{x}) = \sum_{k=1}^K -y_k \log \hat{y}_k$$



今日の学習内容

- 回帰問題の解法を学びます
- 線形回帰モデル(入出力間の関係として線形関数を使った回帰モデル)を学びます
 - 線形回帰：最小二乗法、リッジ回帰、正規方程式



回帰モデル: 学習

- 回帰モデルのための教師データの例 (学習時)

乗客数	離陸重量	燃料重量比	巡行燃費	航続距離
120	50.0	0.250	0.65	3850
108	60.6	0.309	0.6	5820
108	65.1	0.312	0.6	6610
496	375.0	0.414	0.55	11800

学習



回帰
モデル

入力
(特徴量ベクトル)

目標出力
(実数値)



回帰モデル：推論

- 予測の例（推論時）

乗客数	離陸重量	燃料重量 比	巡行燃費	航続距離
48	77.6	0.412	0.6	不明

入力
(特徴量ベクトル)

学習済み

回帰
モデル

出力
(実数値)

予測する



線形回帰モデル

- 入力ベクトル x から出力 y を得る関数が x の線形関数(w と x の内積)

$$y = w_0 + w_1x_1 + w_2x_2 + \cdots + w_Kx_K = \sum_{i=0}^K w_ix_i = \langle w, x \rangle = w^T x$$

$$\text{ただし、} x = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_K \end{pmatrix}, w = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_K \end{pmatrix}$$

T は転置

(w_0, w_1, \dots, w_K) のこと

※ x が1次元の時 : $y = w_1x_1 + w_0$



線形回帰モデルの学習

- 学習: 損失の最小化

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} L(\mathbf{w})$$

$L(\mathbf{w})$: データ全体の損失

- 線形回帰モデルの損失関数

- 最小二乗法

$$L(\mathbf{w}) = \sum_{(x,y) \in D} \underbrace{(y}_{\text{正解出力}} - \underbrace{\mathbf{w}^T \mathbf{x}}_{\text{予測出力}})^2$$

- リッジ回帰

$$L(\mathbf{w}) = \sum_{(x,y) \in D} (y - \mathbf{w}^T \mathbf{x})^2 + \underbrace{\frac{\lambda}{2} \|\mathbf{w}\|^2}_{\text{正則化項}}$$

$$\|\mathbf{w}\|^2 = \left(\sqrt{w_0^2 + w_1^2 + \cdots + w_K^2} \right)^2 = w_0^2 + w_1^2 + \cdots + w_K^2$$



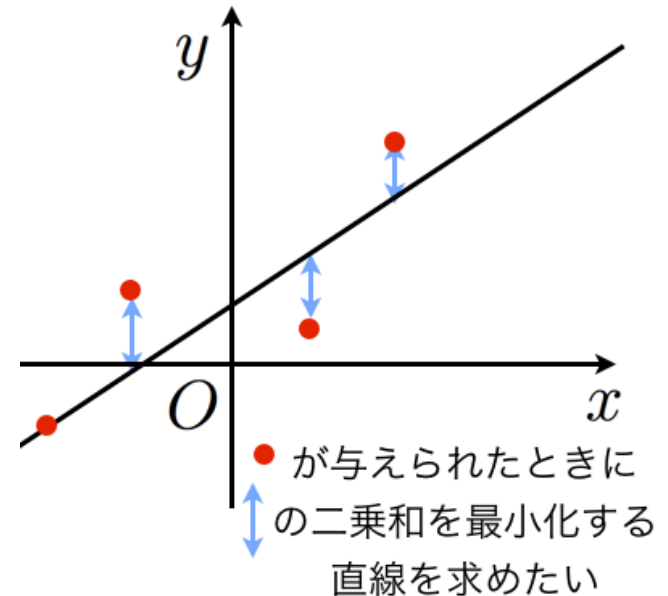
最小二乗法

$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}} \sum_{(x,y) \in D} (y - \mathbf{w}^T \mathbf{x})^2$$

教師データDにおける入出力（ x と y ）の関係を
表すもっともらしい関数を求める

各教師データとの**差の二乗和**
を最小にする関数

例： x が1次元（モデルが直線）の時



回帰モデルの学習例

- 教師データ $D = \{(1,2), (2,3), (3,5)\}$ から最小二乗法により学習される直線の式 $f(x) = ax + b$ を求めなさい。

$$L(a, b) = \sum_{(x,y) \in D} (y - f(x))^2$$



回帰モデルの学習

- 教師データ $D = \{(1,2), (2,3), (3,5)\}$ から最小二乗法により学習される直線の式 $f(x) = ax + b$ を求めなさい。

$$L(a, b) = \sum_{(x,y) \in D} (y - f(x))^2$$

$$L(a, b) = (2 - a - b)^2 + (3 - 2a - b)^2 + (5 - 3a - b)^2$$

$$\begin{aligned} \frac{\partial L(a, b)}{\partial a} &= -2(2 - a - b) - 4(3 - 2a - b) - 6(5 - 3a - b) = \\ &= -46 + 28a + 12b = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial L(a, b)}{\partial b} &= -2(2 - a - b) - 2(3 - 2a - b) - 2(5 - 3a - b) = \\ &= -20 + 12a + 6b = 0 \end{aligned}$$

解くと、 $a = \frac{3}{2}, b = \frac{1}{3}$ 。従って、 $f(x) = \frac{3}{2}x + \frac{1}{3}$ 。

$\hat{\mathbf{w}}$ を求めると...

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{(\mathbf{x}, y) \in D} (y - \mathbf{w}^T \mathbf{x})^2 = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{y} - X\mathbf{w}\|^2$$

ただし、 $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_K \end{pmatrix}$, $X = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ \vdots & & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NK} \end{pmatrix}$

$$\underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{y} - X\mathbf{w}\|^2 = \underset{\mathbf{w}}{\operatorname{argmin}} (\mathbf{y} - X\mathbf{w})^T (\mathbf{y} - X\mathbf{w})$$

$$\frac{\partial}{\partial \mathbf{w}} (\mathbf{y} - X\mathbf{w})^T (\mathbf{y} - X\mathbf{w}) = 0 \text{ を解けばよい}$$

解くと... $(X^T X)\mathbf{w} = X^T \mathbf{y}$ **正規方程式**

$$\hat{\mathbf{w}} = (X^T X)^{-1} X^T \mathbf{y}$$



【参考】 正規方程式の導出

$$\frac{\partial}{\partial \mathbf{w}} (\mathbf{y} - X\mathbf{w})^T (\mathbf{y} - X\mathbf{w}) = 0$$

$$\frac{\partial}{\partial \mathbf{w}} (\mathbf{y}^T \mathbf{y} - \mathbf{w}^T X^T \mathbf{y} - \mathbf{y}^T X \mathbf{w} + \mathbf{w}^T X^T X \mathbf{w}) = 0$$

左辺の各項目の \mathbf{w} による偏微分

転置の公式 : $(AB)^T = B^T A^T$

内積の微分の公式 : $\frac{\partial}{\partial \mathbf{w}} \mathbf{w}^T \mathbf{a} = \frac{\partial}{\partial \mathbf{w}} \mathbf{a}^T \mathbf{w} = \mathbf{a}$ (\mathbf{a} は定数)

2次形式の微分の公式 : $\frac{\partial}{\partial \mathbf{w}} \mathbf{w}^T A \mathbf{w} = (A + A^T) \mathbf{w}$

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{y}^T \mathbf{y} = 0, \quad \frac{\partial}{\partial \mathbf{w}} \mathbf{w}^T X^T \mathbf{y} = X^T \mathbf{y}, \quad \frac{\partial}{\partial \mathbf{w}} \mathbf{y}^T X \mathbf{w} = [\mathbf{y}^T X]^T = X^T \mathbf{y}$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{w}^T X^T X \mathbf{w} = (X^T X + (X^T X)^T) \mathbf{w} = 2X^T X \mathbf{w}$$

$$-2X^T \mathbf{y} + 2X^T X \mathbf{w} = 0 \text{ まとめると、} (X^T X) \mathbf{w} = X^T \mathbf{y}$$



x が1次元（直線）の場合

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

$$X^T X = \begin{bmatrix} N & \sum_{n=1}^N x_n \\ \sum_{n=1}^N x_n & \sum_{n=1}^N x_n^2 \end{bmatrix}, \quad (X^T X)^{-1} = \frac{1}{N \sum_{n=1}^N x_n^2 - (\sum_{n=1}^N x_n)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix},$$

$$X^T \mathbf{y} = \begin{bmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N x_n y_n \end{bmatrix}.$$

なので、

$$\hat{w}_0 = \frac{1}{N} \left(\sum_{n=1}^N y_n - \hat{w}_1 \sum_{n=1}^N x_n \right),$$

$$\hat{w}_1 = \frac{N \sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n \cdot \sum_{n=1}^N y_n}{N \sum_{n=1}^N x_n^2 - (\sum_{n=1}^N x_n)^2}.$$



演習

- 教師データ= $\{(1,2), (2,3), (3,5)\}$ から最小二乗法により学習される直線の式を求めなさい。ただし、正規方程式の解を用いて答えなさい。



演習（答え）

- 教師データ= {(1,2),(2,3),(3,5)} から最小二乗法により学習される直線の式を求めなさい。

$$N = 3, \quad \sum_{n=1}^N x_n = 6, \quad \sum_{n=1}^N y_n = 10, \quad \sum_{n=1}^N x_n y_n = 23, \quad \sum_{n=1}^N x_n^2 = 14$$

$$\hat{w}_1 = \frac{N \sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n \cdot \sum_{n=1}^N y_n}{N \sum_{n=1}^N x_n^2 - (\sum_{n=1}^N x_n)^2} = \frac{3 \times 23 - 6 \times 10}{3 \times 14 - 36} = \frac{9}{6} = \frac{3}{2}$$

$$\hat{w}_0 = \frac{1}{N} \left(\sum_{n=1}^N y_n - \hat{w}_1 \sum_{n=1}^N x_n \right) = \frac{1}{3} \left(10 - \frac{3}{2} \times 6 \right) = \frac{1}{3}$$

$$\text{答え} \quad \underline{\underline{\frac{3}{2}x + \frac{1}{3}}}$$



リッジ回帰

- 最小二乗法に2次の正則化項を導入することで過学習を緩和

※過学習: 訓練データに過剰に適合してしまうこと

$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}} \left\{ \sum_{(x,y) \in D} (y - \mathbf{w}^T \mathbf{x})^2 + \underline{\frac{\lambda}{2} \|\mathbf{w}\|^2} \right\}$$

正則化項

解くと



$$(X^T X + \lambda I) \mathbf{w} = X^T \mathbf{y}$$

$$\hat{\mathbf{w}} = (X^T X + \lambda I)^{-1} X^T \mathbf{y}$$

$$\|\mathbf{w}\|^2 = \sqrt{w_0^2 + w_1^2 + \cdots + w_K^2}$$

※ λ は正則化定数
(予め人手で定める)

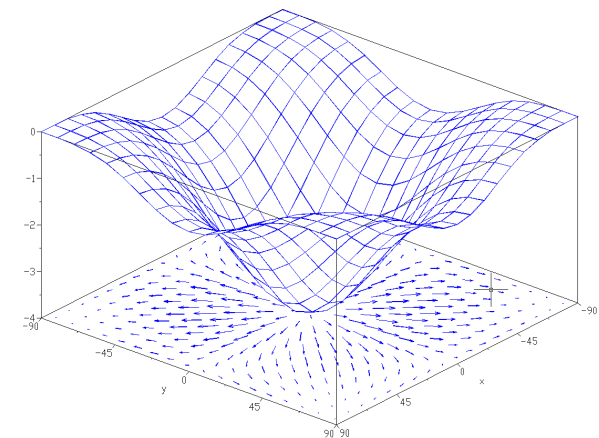
※ I は単位行列



最適化に関して

- 逆行列が簡単に求まる場合: 正規方程式を利用
- 逆行列演算が難しい場合: 数値計算法を利用
 - ニュートン法、準ニュートン法
 - 最急降下法
 - オンライン学習 (確率的勾配法(SGD), AdaGrad, Adamなど)
- 勾配を用いて計算する手法が多い

勾配 $\nabla L = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_m} \right)$



© Simiprof CC 表示 3.0



まとめ

- **線形回帰モデル**

- 最小二乗法
- リッジ回帰
- 正規方程式

- **最適化**

- 正規方程式による最適解の導出
- 逆行列演算が難しい場合は勾配に基づく数値計算法

