神経科学

2023 年 10 月 12 日

目次

第 I 部
機械学習入門

Day1

1 線形回帰モデル

機械学習手法のうち最も基本的なモデルである線形回帰モデルについて述べる．

1.1 線形回帰モデル

線形回帰とは，入力 $\vec{x}$ と出力 y の間に線形的な関係があると仮定し，訓練データ集合 $\mathcal{D} = \{(\vec{x}_1, y_1),\ (\vec{x}_2, y_2),\ (\vec{x}_N, y_N)\}$ から，線形モデル

$$y_i = \vec{w}^t \vec{x}_i = w_1(\vec{x}_i)_1 + w_2(\vec{x}_i)_2 + \cdots + w_d(\vec{x}_i)_d \tag{1}$$

の未知パラメータ $\vec{w}$ を推定し，未知入力に対する予測を行うモデルである．ここで，入力 $\vec{x}$ は d 次元ベクトル，出力はスカラーである．式中の表記 $(\vec{x}_i)_j$ は，訓練データ集合 i 番目の入力ベクトルの j 番目の要素である．これは出力 y_i を基底 x で展開し，その展開係数を推定することが目的であるとも言える．

1.2 特徴量を入れた線形回帰モデル

入力として，直接訓練集合の入力ベクトル $\vec{x}$ を入力するのではなく，入力ベクトル $\vec{x}$ を特徴量ベクトル $\vec{\phi}(\vec{x})$ へ変換し，特徴量ベクトルに対する線形モデル

$$y_i = \vec{w}^t \vec{\phi}(\vec{x}_i) \tag{2}$$

を考えることで，非線形的な関係性を表現することができる．入力ベクトルが非線形変換されているが，パラメータ $\vec{w}$ に対して，線形になっていれば，線形回帰モデルである．

■**例** 1: 入力と出力がともにスカラーの場合を考えよう．入力 x に対する出力 y の関係を多項式によってフィッティングする場合を考える；

$$y = \sum_{k=0}^{d} w_k x^k \tag{3}$$

ここで，x^k は x の k 乗を表し，$x^0 = 1$, $w_0 = b$ である．これは，入力が 1 次元であったが特徴量ベクトルに変換されたことで，次元が d に増えていることに注意せよ．このとき特徴量ベクトルは次のようにかける：

$$\vec{\phi}(x) = \begin{pmatrix} 1 \\ \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_d(x) \end{pmatrix} = \begin{pmatrix} 1 \\ x^1 \\ x^2 \\ \vdots \\ x^d \end{pmatrix} \tag{4}$$

訓練データの入力，出力ベクトルの次元がともに d 次元のときを考える．このとき，入力ベクトルと出力ベクトルはそれぞれ $\vec{x} = (x_1\ x_2\ x_3\ \cdots x_d)$, $\vec{y} = (y_1\ y_2\ y_3\ \cdots y_d)$ である．このとき，特徴量ベクトルへの変換は,

$$\begin{aligned} \vec{y}^{\ t} &= \vec{w}^{\ t}\hat{\phi} = \vec{w}^{\ t}(\vec{\phi}(x_1)\ \vec{\phi}(x_2)\ \cdots\ \vec{\phi}(x_d)) \\ &= (w_1\ w_2\ \cdots\ w_d) \begin{pmatrix} \phi_0(x_1) & \phi_0(x_2) & \cdots & \phi_0(x_d) \\ \phi_1(x_1) & \phi_1(x_2) & \cdots & \phi_1(x_d) \\ \phi_2(x_1) & \phi_2(x_2) & \cdots & \phi_2(x_d) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_n(x_1) & \phi_n(x_2) & \cdots & \phi_n(x_d) \end{pmatrix} \end{aligned} \tag{5}$$

1.3 多次元の線形回帰モデル

高次元の線形回帰モデル

$$\vec{y}_i = {}^t\vec{w}\vec{x}_i \tag{6}$$

を考える．このモデルに対して，訓練データ集合 $\mathcal{D} = \{(\vec{x}_1, \vec{y}_1), (\vec{x}_2, \vec{y}_2), \cdots, (\vec{x}_N, \vec{y}_N)\}$ が与えられたとき，予測誤差の二乗和

$$L(\vec{w}) = \sum_{i=1}^{N} \|\vec{y}_i - {}^t\vec{w}\vec{x}_i\|^2 \tag{7}$$

を損失関数として，損失関数が最小となるようなパラメータ $\vec{w}^*$ を求めれば良い．これは数式で書けば

$$\vec{w}^* = \arg\min_{\vec{w}} L(\vec{w}) \tag{8}$$

である．最小二乗法による最適パラメータ $\vec{w}^*$ は，訓練データ集合から行列

$$\hat{X} = \begin{pmatrix} {}^t\vec{x}_1 \\ {}^t\vec{x}_2 \\ {}^t\vec{x}_3 \\ \vdots \\ {}^t\vec{x}_N \end{pmatrix} = \begin{pmatrix} (\vec{x}_1)_1 & (\vec{x}_1)_2 & \cdots & (\vec{x}_1)_d \\ (\vec{x}_2)_1 & (\vec{x}_2)_2 & \cdots & (\vec{x}_2)_d \\ (\vec{x}_3)_1 & (\vec{x}_3)_2 & \cdots & (\vec{x}_3)_d \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}_N)_1 & (\vec{x}_N)_2 & \cdots & (\vec{x}_N)_d \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} \tag{9}$$

を構成したとき，方程式

$${}^t\hat{X}\hat{X}\vec{w}^* = {}^t\hat{X}\vec{y} \tag{10}$$

を満たす．この方程式を正規方程式という．この方程式は ${}^t\hat{X}\hat{X}$ が正規行列，つまり逆行列が存在すれば $\vec{w}^* = ({}^t\hat{X}\hat{X})^{-1}\ {}^t\hat{X}\vec{y}$ と解ける．正規方程式の導出を次に示す．

1.4　具体的な実装・結果

第 II 部
Boltzmann machine

　線形回帰とは，入力 $\vec{x}$ と出力 y の間に線形的な関係があると仮定し，訓練データ集合 $\mathcal{D} = \{(\vec{x}_1, y_1),\ (\vec{x}_2, y_2),\ (\vec{x}_N, y_N)\}$

2 Boltzmann machine

3 可視変数のみのボルツマンマシン

4 隠れ変数を持つボルツマンマシン

ボルツマン機械学習では，統計力学で基本となるカノニカル分布 (ボルツマン分布) に従ってデータ $\vec{x}$ が生成されると考える：

$$P_B(\vec{x}|\vec{\Theta}) = \frac{1}{Z_B(\vec{\Theta})} \exp -E(\vec{x}|\vec{\Theta}) \tag{11}$$

ここで，確率分布の引数 $\vec{x}$, $\vec{\Theta}$ はそれぞれイジング変数とパラメータを表す．パラメータ $\vec{\Theta}$ の中身は後で言及を行う．ここで，エネルギー関数は

$$E(\vec{x}|\vec{\gamma},\vec{c}) = -\sum_{(i,j)\in E} \gamma_{i,j} x_i x_j - \sum_{i\in V} c_i x_i \tag{12}$$

のように相互作用項と外場項（バイアス項）の和により定義される．$Z_B(\vec{\Theta})$ は規格化定数であり，分配関数と呼ばれ，次のように定義される：

$$Z_B(\vec{\Theta}) = \sum_{\vec{x}\in\{\pm1\}^n} \exp\left(-E(\vec{x}|\vec{\Theta})\right) \tag{13}$$

ここで，上式中の和の記号は,

$$\sum_{\vec{x}} \equiv \prod_{i\in E} \sum_{x_i\in\{\pm1\}} = \sum_{x_1\in\{\pm1\}} \sum_{x_2\in\{\pm1\}} \sum_{x_3\in\{\pm1\}} \cdots \sum_{x_{|E|}\in\{\pm1\}} \tag{14}$$

$$\begin{aligned} E(\vec{v},\vec{h}) = &-\sum_{i,j} w_{i,j} v_i h_j - \sum_{j,j'} \alpha_{j,j'} h_j h_{j'} - \sum_{i,i'} \beta_{i,i'} v_i v_{i'} \\ &- \sum_{j} b_{j,j'} h_j - \sum_{i} a_i v_i \end{aligned} \tag{15}$$

これを行列で表すと以下のようになる：

$$E(\vec{v},\vec{h}) = -\left[\vec{v}^t \hat{W} \vec{h} + \vec{h}^t \hat{A} \vec{h} + \vec{v}^t \hat{B} \vec{h} + \vec{a}^t \vec{h} + \vec{b}^t \vec{h},\right] \tag{16}$$

ここで，式中のベクトルと行列はそれぞれである：

$$\hat{A} = \begin{pmatrix} (\vec{x}_1)_1 & (\vec{x}_1)_2 & \cdots & (\vec{x}_1)_d \\ (\vec{x}_2)_1 & (\vec{x}_2)_2 & \cdots & (\vec{x}_2)_d \\ (\vec{x}_3)_1 & (\vec{x}_3)_2 & \cdots & (\vec{x}_3)_d \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}_N)_1 & (\vec{x}_N)_2 & \cdots & (\vec{x}_N)_d \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} \tag{17}$$

5　RBM

制限ボルツマンマシンは

$$\alpha_{j,j'} = \beta_{i,i'} = 0 \tag{18}$$

すなわち，

$$\vec{a}^t\vec{h} = \vec{b}^t\vec{h} = 0 \tag{19}$$

としたものをいう．これは，可視層同士，隠れ層同士の結合を考えないモデルに帰着する．制限ボルツマンマシンのエネルギー関数は

$$E(\vec{v}, \vec{h}) = -\sum_{i,j} w_{i,j} v_i h_j - \sum_j b_j h_j - \sum_i a_i v_i \tag{20}$$

$$= -\vec{v}^t \hat{W} \vec{h} - \vec{h}^t \hat{A} \vec{h} - \vec{v}^t \hat{B} \vec{h} \tag{21}$$

これを行列で表すと以下のようになる：

5.1　制限ボルツマンマシンの条件付き確率の独立性

可視層を固定したもとでの，制限ボルツマンマシンの隠れ層の条件付き確率は，以下のように書ける：

$$P(\vec{h}|\vec{v}; \vec{\Theta}) = \frac{P(\vec{h}, \vec{v}; \vec{\Theta})}{P(\vec{v}|\vec{\Theta})} = \prod_{i=1}^{N} \frac{\exp \lambda_j^H h_j}{2 \cosh \lambda_j^H h_j} \tag{22}$$

ここで，

$$\lambda_i^H \equiv b_i + \sum_{j=1}^{N} w_{i,j} v_j. \tag{23}$$

また，隠れ層を固定したもとでの，制限ボルツマンマシンの可視層の条件付き確率は，以下のように書ける：

$$P(\vec{v}|\vec{h};\vec{\Theta}) = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} = \prod_{i=1}^{N} \frac{\exp \lambda_j^V v_j}{2\cosh \lambda_j^V v_j} \tag{24}$$

ここで，

$$\lambda_i^V \equiv a_i + \sum_{j=1}^{N} w_{i,j} h_j. \tag{25}$$

これを証明する：

$$P(\vec{v}|\vec{h};\vec{\Theta}) = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} \tag{26}$$

を考える．可視変数 $\vec{v}$ に関する周辺確率は

$$P(\vec{h}|\vec{\Theta}) = \sum_{\vec{v}} P(\vec{v},\vec{h};\vec{\Theta}) \tag{27}$$

と書ける．これを用いることで，

$$\begin{aligned} P(\vec{v}|\vec{h};\vec{\Theta}) &= \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{\sum_{\vec{v}} P(\vec{v},\vec{h};\vec{\Theta})} \\ &= \frac{\exp\left[\sum_{i,j} w_{i,j} v_i h_j + \sum_i a_i v_i + \sum_j b_j h_j\right]}{\sum_{\vec{v}} \exp\left[\sum_{i,j} w_{i,j} v_i h_j + \sum_i a_i v_i + \sum_j b_j h_j\right]} \\ &= \frac{\exp\left[\sum_i v_i \left\{\sum_{i,j} w_j h_j + a_i\right\} + \sum_j b_j h_j\right]}{\sum_{\vec{v}} \exp\left[\sum_i v_i \left\{\sum_j w_{i,j} h_j + a_i\right\} + \sum_j b_j h_j\right]} \end{aligned} \tag{28}$$

ここで，$\vec{v}$ に対する和に関係のない項 $+\sum_j b_j h_j$ は約分できる：

$$
\begin{aligned}
P(\vec{v}|\vec{h};\vec{\Theta}) &= \frac{\exp\left[\sum_i v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]\exp\left[+\sum_j b_j h_j\right]}{\sum_{\vec{v}}\exp\left[\sum_i v_i\left\{\sum_{i,j} w_j h_j + a_i\right\}\right]\exp\left[+\sum_j b_j h_j\right]} \\
&= \frac{\exp\left[\sum_i v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]}{\sum_{\vec{v}}\exp\left[\sum_i v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]} \\
&= \frac{\prod_{i=1}^N \exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]}{\sum_{\vec{v}}\prod_{i=1}^N \exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]} \\
&= \frac{\prod_{i=1}^N \exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]}{\sum_{\vec{v}}\prod_{i=1}^N \exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]}
\end{aligned} \tag{29}
$$

ここで，

$$
\begin{aligned}
\sum_{\vec{v}}\prod_{i=1}^N \exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right] &= \prod_{i=1}^N\left[\sum_{v_i}\exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]\right] \\
&= \prod_{i=1}^N 2\cosh\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]
\end{aligned} \tag{30}
$$

$$
P(\vec{v}|\vec{h};\vec{\Theta}) = \prod_{i=1}^N \frac{\exp\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]}{2\cosh\left[v_i\left\{\sum_j w_{i,j}h_j + a_i\right\}\right]} \tag{31}
$$

6 neural quantum state

7 重要な公式集

$$\sum_{x=\pm 1} \exp(ax) = \exp a + \exp -a = 2\cosh a \tag{32}$$

多変数への拡張

$$\begin{aligned}
\sum_{\vec{x}\in\{\pm 1\}^N} \exp(\vec{a}\ {}^t\vec{x}) &= \sum_{\vec{x}\in\{\pm 1\}^N} \exp\Big(\sum_{i=1}^{N} a_i x_i\Big) \\
&= \sum_{\vec{x}\in\{\pm 1\}^N} \prod_{i=1}^{N} \exp\Big(a_i x_i\Big) \\
&= \prod_{i=1}^{N} \Bigg[\sum_{x_i=\{\pm 1\}} \exp\Big(a_i x_i\Big)\Bigg] \\
&= \prod_{i=1}^{N} \Bigg[2\cosh(a_i x_i)\Bigg]
\end{aligned} \tag{33}$$

ここで，2 番目から 3 番目の等式に移る際に，

$$\sum_{x_1=\pm 1} \sum_{x_2=\pm 1} \cdots \sum_{x_N=\pm 1} \prod_{i=1}^{N} \exp\Big(a_i x_i\Big) = \prod_{i=1}^{N} \Bigg[\sum_{x_i=\{\pm 1\}} \exp\Big(a_i x_i\Big)\Bigg]$$

という関係式が一般的に成り立つことを用いた．これは，2 次元の場合に簡単に確認で

きる：

$$\sum_{x_1=\pm 1}\sum_{x_2=\pm 1}\prod_{i=1}^{2}\exp\Big(a_i x_i\Big) = \sum_{x_1=\pm 1}\sum_{x_2=\pm 1}\exp\Big(a_1 x_1 + a_2 x_2\Big)$$
$$= e^{a_1+a_2} + e^{a_1-a_2} + e^{-a_1+a_2} + e^{-a_1-a_2}$$
$$= e^{a_1}(e^{a_2} + e^{-a_2}) + e^{-a_1}(e^{a_2} + e^{-a_2})$$
$$= (e^{a_1} + e^{-a_1})(e^{a_2} + e^{-a_2})$$
$$= \prod_{i=1}^{2}\Bigg[\sum_{x_i=\{\pm 1\}}\exp\Big(a_i x_i\Big)\Bigg] \tag{34}$$

8　量子ボルツマン機械学習

参考文献