原子力研究開発機構夏季実習生 repor

2023 年 10 月 12 日

目次

第 I 部
確率論の基礎

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 $\mathcal{D} = \{(\vec{x}_1, y_1), \ (\vec{x}_2, y_2), \ (\vec{x}_N, y_N)\}$

1 Boltzmann machine

2 可視変数のみのボルツマンマシン

3 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：

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

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

$$E(\vec{x}|\vec{\gamma}, \vec{c}) = -\sum_{(i,j)\in E} \gamma_{i,j} x_i x_j - \sum_{i\in V} c_i x_i \tag{2}$$

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：

$$Z_B(\vec{\Theta}) = \sum_{\vec{x}\in\{\pm1\}^n} \exp\left(-E(\vec{x}|\vec{\Theta})\right) \tag{3}$$

ここで，上式中の和の記号は,

$$\sum_{\vec{x}} \equiv \prod_{i\in E} \sum_{x_i\in\{\pm1\}} = \sum_{x_1\in\{\pm1\}} \sum_{x_2\in\{\pm1\}} \sum_{x_3\in\{\pm1\}} \cdots \sum_{x_|E|\in\{\pm1\}} \tag{4}$$

$$E(\vec{v},\vec{h}) = -\sum_{i,j} w_{i,j} v_i h_j - \sum_{j,j'} \alpha_{j,j'} h_j h_{j'} - \sum_{i,i'} \beta_{i,i'} v_i v_{i'}$$

$$- \sum_j b_{j,j'} h_j - \sum_i a_i v_i \tag{5}$$

これを行列で表すと以下のようになる：

$$E(\vec{v},\vec{h}) = -\Big[\vec{v}^t \hat{W} \vec{h} + \vec{h}^t \hat{A} \vec{h} + \vec{v}^t \hat{B} \vec{h} + \vec{a}^t \vec{h} + \vec{b}^t \vec{h},\Big] \tag{6}$$

ここで，式中のベクトルと行列はそれぞれである：

$$\hat{A} = \begin{pmatrix} (\vec{x}_1)_1 & (\vec{x}_1)_2 & \cdots & (\vec{x}_1)_d \\ (\vec{x}_2)_1 & (\vec{x}_2)_2 & \cdots & (\vec{x}_2)_d \\ (\vec{x}_3)_1 & (\vec{x}_3)_2 & \cdots & (\vec{x}_3)_d \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}_N)_1 & (\vec{x}_N)_2 & \cdots & (\vec{x}_N)_d \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} \tag{7}$$

4　RBM

制限ボルツマンマシンは

$$\alpha_{j,j'} = \beta_{i,i'} = 0 \tag{8}$$

すなわち，

$$\vec{a}^t \vec{h} = \vec{b}^t \vec{h} = 0 \tag{9}$$

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

$$E(\vec{v},\vec{h}) = -\sum_{i,j} w_{i,j} v_i h_j - \sum_j b_j h_j - \sum_i a_i v_i \tag{10}$$

$$= -\vec{v}^t \hat{W} \vec{h} - \vec{h}^t \hat{A} \vec{h} - \vec{v}^t \hat{B} \vec{h} \tag{11}$$

これを行列で表すと以下のようになる：

4.1 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：

$$P(\vec{h}|\vec{v};\vec{\Theta}) = \frac{P(\vec{h},\vec{v};\vec{\Theta})}{P(\vec{v}|\vec{\Theta})} = \prod_{i=1}^{N} \frac{\exp \lambda_j^H h_j}{2\cosh \lambda_j^H h_j} \tag{12}$$

ここで,

$$\lambda_i^H \equiv b_i + \sum_{j=1}^{N} w_{i,j} v_j. \tag{13}$$

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：

$$P(\vec{v}|\vec{h};\vec{\Theta}) = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} = \prod_{i=1}^{N} \frac{\exp \lambda_j^V v_j}{2\cosh \lambda_j^V v_j} \tag{14}$$

ここで,

$$\lambda_i^V \equiv a_i + \sum_{j=1}^{N} w_{i,j} h_j. \tag{15}$$

これを証明する：

$$P(\vec{v}|\vec{h};\vec{\Theta}) = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} \tag{16}$$

を考える．可視変数 $\vec{v}$ に関する周辺確率は

$$P(\vec{h}|\vec{\Theta}) = \sum_{\vec{v}} P(\vec{v},\vec{h};\vec{\Theta}) \tag{17}$$

と書ける．これを用いることで，

$$
\begin{aligned}
P(\vec{v}|\vec{h};\vec{\Theta}) &= \frac{P(\vec{v},\vec{h};\vec{\Theta})}{P(\vec{h}|\vec{\Theta})} = \frac{P(\vec{v},\vec{h};\vec{\Theta})}{\sum_{\vec{v}} P(\vec{v},\vec{h};\vec{\Theta})} \\
&= \frac{\exp\left[\sum_{i,j} w_{i,j} v_i h_j + \sum_i a_i v_i + \sum_j b_j h_j\right]}{\sum_{\vec{v}} \exp\left[\sum_{i,j} w_{i,j} v_i h_j + \sum_i a_i v_i + \sum_j b_j h_j\right]} \\
&= \frac{\exp\left[\sum_i v_i \left\{\sum_{i,j} w_j h_j + a_i\right\} + \sum_j b_j h_j\right]}{\sum_{\vec{v}} \exp\left[\sum_i v_i \left\{\sum_j w_{i,j} h_j + a_i\right\} + \sum_j b_j h_j\right]}
\end{aligned} \tag{18}
$$

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：

$$
\begin{aligned}
P(\vec{v}|\vec{h};\vec{\Theta}) &= \frac{\exp\left[\sum_i v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right] \exp\left[+\sum_j b_j h_j\right]}{\sum_{\vec{v}} \exp\left[\sum_i v_i \left\{\sum_{i,j} w_j h_j + a_i\right\}\right] \exp\left[+\sum_j b_j h_j\right]} \\
&= \frac{\exp\left[\sum_i v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]}{\sum_{\vec{v}} \exp\left[\sum_i v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]} \\
&= \frac{\prod_{i=1}^{N} \exp\left[v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]}{\sum_{\vec{v}} \prod_{i=1}^{N} \exp\left[v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]} \\
&= \frac{\prod_{i=1}^{N} \exp\left[v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]}{\sum_{\vec{v}} \prod_{i=1}^{N} \exp\left[v_i \left\{\sum_j w_{i,j} h_j + a_i\right\}\right]}
\end{aligned} \tag{19}
$$

ここで，

$$\sum_{\vec{v}} \prod_{i=1}^{N} \exp\left[v_i \Big\{ \sum_j w_{i,j} h_j + a_i \Big\} \right] = \prod_{i=1}^{N} \left[\sum_{v_i} \exp\left[v_i \Big\{ \sum_j w_{i,j} h_j + a_i \Big\} \right] \right]$$
$$= \prod_{i=1}^{N} 2\cosh\left[v_i \Big\{ \sum_j w_{i,j} h_j + a_i \Big\} \right] \tag{20}$$

$$P(\vec{v}|\vec{h};\vec{\Theta}) = \prod_{i=1}^{N} \frac{\exp\left[v_i \Big\{ \sum_j w_{i,j} h_j + a_i \Big\} \right]}{2\cosh\left[v_i \Big\{ \sum_j w_{i,j} h_j + a_i \Big\} \right]} \tag{21}$$

5 neural quantum state

6 重要な公式集

$$\sum_{x=\pm 1} \exp(ax) = \exp a + \exp -a = 2\cosh a \tag{22}$$

多変数への拡張

$$\sum_{\vec{x} \in \{\pm 1\}^N} \exp(\vec{a}^{\ t} \vec{x}) = \sum_{\vec{x} \in \{\pm 1\}^N} \exp\Big(\sum_{i=1}^{N} a_i x_i \Big)$$
$$= \sum_{\vec{x} \in \{\pm 1\}^N} \prod_{i=1}^{N} \exp\Big(a_i x_i \Big)$$
$$= \prod_{i=1}^{N} \left[\sum_{x_i = \{\pm 1\}} \exp\Big(a_i x_i \Big) \right]$$
$$= \prod_{i=1}^{N} \Big[2\cosh(a_i x_i) \Big] \tag{23}$$

ここで，2 番目から 3 番目の等式に移る際に，

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

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：

$$
\begin{aligned}
\sum_{x_1=\pm 1}\sum_{x_2=\pm 1}\prod_{i=1}^{2}\exp\Big(a_i x_i\Big) &= \sum_{x_1=\pm 1}\sum_{x_2=\pm 1}\exp\Big(a_1x_1+a_2x_2\Big) \\
&= e^{a_1+a_2}+e^{a_1-a_2}+e^{-a_1+a_2}+e^{-a_1-a_2} \\
&= e^{a_1}(e^{a_2}+e^{-a_2})+e^{-a_1}(e^{a_2}+e^{-a_2}) \\
&= (e^{a_1}+e^{-a_1})(e^{a_2}+e^{-a_2}) \\
&= \prod_{i=1}^{2}\Bigg[\sum_{x_i=\{\pm 1\}}\exp\Big(a_i x_i\Big)\Bigg] \qquad (24)
\end{aligned}
$$

第 II 部
標本分布

7　母集団とランダム標本

サイコロを振った場合の 1 回ごとの試行結果や，ねじを抽出した場合のそれぞれのねじのように，実験や観測を行う１つ１つの対象を個体といい，考えている個体全部の集合を母集団という．特に，母集団の中にある個体の数が有限の場合を有限母集団，無限の場合を無限母集団という．

n 個の確率変数 $\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ が互いに独立でそれぞれ同一の確率分布に従うとする．このとき，確率変数 $\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ は，その確率分布がもつ母集団からの大きさ n のランダム標本という．

統計量

確率変数 $\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ をランダム標本とすると，$\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ の関数を統計量という．そして統計量の確率分布を標本分布という．ここで，統計量はランダム変数を変数変換したものであるため，統計量自身も確率変数であることに注意が必要である．

7.1　標本積率

ランダム標本 $\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ に対して，

$$\hat{M}_r = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_i^r \tag{25}$$

を r 次の標本積率という．

特に，$r = 1$ の場合

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_i \tag{26}$$

を標本平均という．また，

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - \bar{X}_i)^r \tag{27}$$

を $\bar{X}_n$ のまわりの r 次の標本積率と呼ぶ．このとき，特に $r = 2$ の場合を標本分散と呼び，

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - \bar{X}_i)^2 \tag{28}$$

と表す．また，

$$\hat{S}_n = \sqrt{\hat{S}_n^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - \bar{X}_i)^2} \tag{29}$$

を標本標準偏差という．

7.2　標本平均の期待値と分散

$\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_n$ が母集団確率変数 $\hat{X}$ からの大きさ n のランダム標本とする．母集団の平均と分散がそれぞれ，

$$\mathbb{E}[\hat{X}] = \mu \tag{30}$$

$$\mathrm{Var}[\hat{X}] = \sigma^2 \tag{31}$$

であるとき，標本平均の期待値と分散は

$$\mathbb{E}[\bar{X}_n] = \mu \tag{32}$$

$$\mathrm{Var}[\bar{X}_n] = \sigma^2/n \tag{33}$$

となる．これを示す：期待値と分散に関する線形性と，ランダム標本の期待値と分散がそれぞれ $\mathbb{E}[\hat{X}_i]=\mu$, $\mathrm{Var}[\bar{X}_n]=\sigma^2$, $(i=1,\ \cdots,\ n)$ に従うことを用いると，

$$\mathbb{E}[\bar{X}_n]=E\left[\sum_{i=1}^{n}\hat{X}_i/n\right]=\frac{1}{n}E\left[\sum_{i=1}^{n}\hat{X}_i\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\hat{X}_i]=\frac{1}{n}\cdot n\mu=\mu \tag{34}$$

$$\mathrm{Var}[\bar{X}_n]=\mathrm{Var}\left[\sum_{i=1}^{n}\hat{X}_i/n\right]=\frac{1}{n}\mathrm{Var}\left[\sum_{i=1}^{n}\hat{X}_i\right]=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}[\hat{X}_i]=\frac{1}{n}\cdot n\sigma^2=\sigma^2 \tag{35}$$

8　静的モンテカルロ法

8.1　1 次元モンテカルロ積分

モンテカルロ積分の目的は，乱数を用いて，積分値を推定することである．区間 $a\le x\le b$ 上のなめらかな関数 $g(x)$ に関する積分 I を考える：

$$I=\int_a^b g(x)dx \tag{36}$$

積分区間 $a\le x\le b$ に対応して，一様分布 $U[a,b]$ に基づいて，乱数を生成し，これを用いて，積分 I を推定することを考える．

確率変数 $\hat{X}$ が一様分布 $U[a,b]$ に従うとする：$\hat{X}\sim U[a,b]$．統計量 $g(\hat{X})$ の期待値 $\mathbb{E}[g(\hat{X})]$ に関して，以下が成り立つ：

$$\mathbb{E}[g(\hat{X})]=\int_a^b g(x)f_{\hat{X}}(x)dx=\frac{1}{(b-a)}\int_a^b g(x)dx \tag{37}$$

ここで，$f_{\hat{X}}(x)$ は一様分布 $U[a,b]$ に対する確率変数 $\hat{X}$ の確率密度関数であり，以下のように定義される：

$$f_{\hat{X}}(x)=\begin{cases}\dfrac{1}{(b-a)} & a\le x\le b\\ 0 \text{ other}\end{cases} \tag{38}$$

次に，一様分布 $U[a,b]$ に従う確率変数 $\hat{X}$ で記述される系と完全に同じ系を独立に N 個用意する：

$$\{\hat{X}_1,\ \hat{X}_2,\ \cdots,\ \hat{X}_N\} \tag{39}$$

そして，統計量 $g(\hat{X}_1)$，$g(\hat{X}_2)$，$\cdots$，$g(\hat{X}_N)$ の標本平均 $\hat{\theta}$ を考える：

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} g(\hat{X}_i) \tag{40}$$

ここで，サンプル数 N を増やしていけば，$(N \to \infty)$, 大数弱の法則より，

$$\lim_{N\to\infty}\hat{\theta} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g(\hat{X}_i) = \mathbb{E}[g(\hat{X})] = \frac{1}{(b-a)}\int_a^b g(x)dx = \frac{1}{(b-a)}I \tag{41}$$

が成立する．これをより詳しく見てみる．上式を変形すれば,

$$I = (b-a)\lim_{N\to\infty}\hat{\theta} = \lim_{N\to\infty}\frac{(b-a)}{N}\sum_{i=1}^{N} g(\hat{X}_i) \tag{42}$$

となる．この式が意味するのは，積分 I に対して，区分求積法を実行することに対応している．ここで，サンプル数 N は積分区間の分割数に対応している．これが，乱数を用いて，積分が実行できる理由である．

実際のアルゴリズムは，以下のように実行される：

Step1

まず，一様分布 $U[a,b]$ にしたがって，乱数を発生させ，確率変数 $\hat{X}_1$，$\hat{X}_2$，$\cdots$，$\hat{X}_N$ にそれぞれ対応する観測値 x_1，x_2，$\cdots$，x_N をサンプリングする．

Step2

そして,$g(\hat{X}_1)$, $g(\hat{X}_2)$, $\cdots$, $g(\hat{X}_N)$ にそれぞれ対応する観測値 $g(x_1)$, $g(x_2)$, $\cdots$, $g(x_N)$ を求め，それらの平均

$$\theta = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \tag{43}$$

を求めることで，積分 I が計算できる．

先ほどまでは，一様分布に従って，乱数を発生させ，積分を実行していた．また，一様分布を使うと，得たい期待値の分散は基本的に大きくなる．したがって，分散の小さい分布を用いてサンプリングを行う必要がある．一様分布以外の確率分布からサンプリングを行うことも可能である．これを実行するのが重点サンプリング (Importance Sampling) と呼ばれる方法である．

重点サンプリングでは，まず，求めたい非積分関数 $g(x)$ をサンプリングしたい分布 $P_{\hat{X}}(x)$ で割った関数 $g(x)/P_{\hat{X}}(x)$ を考える．ここで，確率変数 $\hat{X}$ はある確率分布 $P_{\hat{X}}(x)$ にしたがっている．このとき，確率分布 $P_{\hat{X}}(x)$ に関する $g(x)/P_{\hat{X}}(x)$ の期待値は

$$\mathbb{E}\left[\frac{g(\hat{X})}{P_{\hat{X}}(x)}\right] = \int \frac{g(x)}{P_{\hat{X}}(x)} P_{\hat{X}}(x)dx = \int g(x)dx \tag{44}$$

となる．そして，確率分布 $P_{\hat{X}}(x)$ に従う確率変数 $\hat{X}$ で記述される系と完全に同じ系を独立に N 個用意する：

$$\{\hat{X}_1,\ \hat{X}_2,\ \cdots,\ \hat{X}_N\} \tag{45}$$

そして，統計量 $g(\hat{X}_1)/P(\hat{X}_1)$，$g(\hat{X}_2)/P(\hat{X}_2)$，$\cdots$，$g(\hat{X}_N)/P(\hat{X}_N)$ の標本平均 $\hat{\theta}$ を考える：

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(\hat{X}_i)}{P(\hat{X}_i)} \tag{46}$$

ここで，サンプル数 N を増やしていけば，$(N \to \infty)$, 大数弱の法則より，

$$\lim_{N\to\infty}\hat{\theta} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{g(\hat{X}_i)}{P(\hat{X}_i)} = \mathbb{E}[g(\hat{X})] = \int g(x)dx = I \tag{47}$$

が成立する．したがって，積分 $\int g(x)dx$ が推定できることがわかる．

8.2 逆関数法を用いた確率分布の生成

9 Ising モデルのモンテカルロシミュレーション

全ての状態についての足し上げは極めて困難である．そこで，あるスピン配位 C が実現する確率を

$$P(C) = \frac{1}{Z}(e^{-\beta E[C]}) \tag{48}$$

に従い，確率的にスピン配位 C を生成して，生成したスピン配位 C を使った平均を用いて期待値を推定する．

しかしスピン配位 C の実現確率 (48) は直接は計算できない．なぜならば，分母にある分配関数 Z を計算するためには，結局すべての組合せを取る必要があるからである．そこで，分配関数 Z を計算を避けてスピン配位を生成する確率分布を用意する必要がある．我々が求める確率分布を生成する手法の一つがマルコフ連鎖モンテカルロ法である．

9.1　基本原理

配位に対応する物理量 $\hat{C}$ が確率分布 $P(C)$ に従うとする．このとき，推定量 $f(\hat{C})$ に関する期待値は一般的に以下のように書ける：

$$\mathbb{E}[f(\hat{C})] = \sum_{C} f(C)P(C) \tag{49}$$

ここで，$f(\hat{C})$ も（配位に関する）物理量であり，磁化率などが対応している．また，$\sum_C$ は全ての可能なスピン配位についての足し上げを表す．格子点数（スピン数）が N 個の Ising モデルを考えるとき，取り得る全配位数は $|C| = 2^N$ 個となる．

ここで，2^N 個の配位についての和は取りたくない．そこで，2^N 個よりも少ない，有限のサンプル数 $|C_{\mathrm{fin}}|$ のサンプル

$$C_{\mathrm{fin}} = \{\hat{C}_1,\ \hat{C}_2,\ \cdots,\ \hat{C}_{|C_{\mathrm{fin}}|}\} \tag{50}$$

から配位を用意する．ここで，サンプル C_{fin} はそれぞれ，$\hat{C}$ と同じ確率分布に従う．すなわち，C_{fin} はそれぞれ $\hat{C}$ と完全に同じ系を表す．そして，期待値 $\mathbb{E}[f(\hat{C})]$ を評価する代わりに，

$$\hat{\theta} = \frac{1}{|C_{\mathrm{fin}}|} \sum_{\hat{C}_k \in C_{\mathrm{fin}}} f(\hat{C}_k) \tag{51}$$

をサンプリングによって評価する．これは，大数弱の法則から，サンプル数を多くとれば，

$$\mathbb{E}[f(\hat{C})] \simeq \frac{1}{|C_{\mathrm{fin}}|} \sum_{C_k \in C_{\mathrm{fin}}} f(C_k) \tag{52}$$

に近づく．この確率分布 $P(C)$ をマルコフ連鎖モンテカルロ法で作り，この分布に従い配位をサンプリングしてやり，平均を計算すれば，物理量を求めることができる．

参考文献