機械学習（9回目）

創域理工学部　情報計算科学科
桂田　浩一

1

11/16/2023

前回の復習

- ボルツマンマシン
  - 対数尤度の偏微分

3

ボルツマンマシン

- エネルギー関数の分布がボルツマン分布を取るニューラルネットワークをボルツマンマシンと呼ぶ.

↓

十分に更新した時の状態が「分布」になるのなら,最初から「分布」を学習しては？

⇩

例えば「あ」と発話したときの音声波形とか

11/16/2023

4

分布が正しく学習できているか？

- 最尤推定になっているかどうかで判断する

最尤推定：
　データ集合 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ が与えられたとき,$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ が未知の確率分布 $P_g(\boldsymbol{x})$ に従って生成されているとする.
　$P_g(\boldsymbol{x})$ を知る方法は無いので,適当な関数 $P(\boldsymbol{x}|\boldsymbol{\theta})$ を用意し,尤度関数 $L(\boldsymbol{\theta}) = \prod_{n=1}^{N} P(\boldsymbol{x}_n|\boldsymbol{\theta})$ が最大になる $\boldsymbol{\theta}$ のとき,$P(\boldsymbol{x}|\boldsymbol{\theta})$ は $P_g(\boldsymbol{x})$ に近いと考える.

データ集合 $\boldsymbol{X}$ から $\boldsymbol{\theta}$ を推定する！！

11/16/2023

5

対数尤度

- $\ln L(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln P(\boldsymbol{x}_n|\boldsymbol{\theta})$

$P(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} e^{-\Phi(\boldsymbol{x},\boldsymbol{\theta})}$ より

$\ln P(\boldsymbol{x}_n|\boldsymbol{\theta}) = -\Phi(\boldsymbol{x}_n, \boldsymbol{\theta}) - \ln z(\boldsymbol{\theta})$

$\ln L(\boldsymbol{\theta}) = \sum_{n=1}^{N}(-\Phi(\boldsymbol{x}_n, \boldsymbol{\theta}) - \ln z(\boldsymbol{\theta}))$

これを最大にする $\boldsymbol{\theta}$ のとき $P(\boldsymbol{x}|\boldsymbol{\theta})$ と $P_g(\boldsymbol{x})$ は近い

このような $\boldsymbol{\theta}$ を求めることが分布の学習になる
⇒勾配法で求める

11/16/2023

6

多層ニューラルネットとボルツマンマシンの比較

	学習の方針	学習後のネットワーク
多層ニューラルネット	誤差関数 E の最小化	入力に対して正しい出力を行う
ボルツマンマシン	対数尤度 $\ln L(\boldsymbol{\theta})$ の最大化	十分に更新した時のユニットの値 x_i の分布が正しい

多層ニューラルネットの重みの更新式： $w \leftarrow w - \eta \frac{\partial E}{\partial w}$

ボルツマンマシンの重みの更新式： $w \leftarrow w + \eta \frac{\partial \ln L(\boldsymbol{\theta})}{\partial w}$

ボルツマンマシンのバイアスの更新式： $b \leftarrow b + \eta \frac{\partial \ln L(\boldsymbol{\theta})}{\partial b}$

11/16/2023

7

対数尤度関数の偏微分

- $$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial w_{j,i}} = \frac{\partial \sum_{n=1}^{N}(-\Phi(\boldsymbol{x}_n,\boldsymbol{\theta}) - \ln z(\boldsymbol{\theta}))}{\partial w_{j,i}}$$
  $$= \sum_{n=1}^{N}\left(-\boxed{\frac{\partial \Phi(\boldsymbol{x}_n,\boldsymbol{\theta})}{\partial w_{j,i}}} \boxed{-\frac{\partial \ln z(\boldsymbol{\theta})}{\partial w_{j,i}}}\right)$$
  $$= \sum_{n=1}^{N}(x_i^n x_j^n - E_{\boldsymbol{\theta}}[x_i x_j])$$
  $$= \sum_{n=1}^{N}(x_i^n x_j^n) - N E_{\boldsymbol{\theta}}[x_i x_j]$$

8

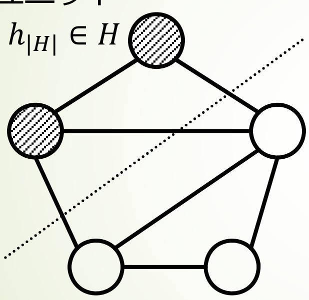
本日の内容

- 制限を加えたボルツマンマシン
  - 隠れユニットのあるボルツマンマシン
  - 制限ボルツマンマシン

9

隠れユニットがあるボルツマンマシン

隠れユニット
$h_1 \cdots h_{|H|} \in H$



可視ユニット
$v_1 \cdots v_{|V|} \in V$

※可視ユニットに隠れユニットを加えることにより，可視ユニット間の関係だけでは表現できない，隠れた関係も考慮した分布を学習できる

※入力は可視ユニットのみに与える

11/16/2023

10

隠れユニットがあるときの尤度関数

- $P(\boldsymbol{x}|\boldsymbol{\theta}) = P(\boldsymbol{v}, \boldsymbol{h}|\boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} e^{-\Phi(\boldsymbol{v}, \boldsymbol{h}, \boldsymbol{\theta})}$

$\Phi(\boldsymbol{x}, \boldsymbol{\theta}) = \Phi(\boldsymbol{v}, \boldsymbol{h}, \boldsymbol{\theta}) = -\sum_{(i,j)} w_{j,i} x_i x_j - \sum_i b_i x_i$

尤度関数：$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} P(\boldsymbol{v}_n, \boldsymbol{h}|\boldsymbol{\theta})$ ？

尤度は与えられたデータと $\boldsymbol{\theta}$ で定義すべき

尤度関数：$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} P(\boldsymbol{v}_n|\boldsymbol{\theta})$

11/16/2023

11

隠れユニットがあるときの尤度関数

- 尤度関数：$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} P(\boldsymbol{v}_n|\boldsymbol{\theta})$

$P(\boldsymbol{x}|\boldsymbol{\theta}) = P(\boldsymbol{v}, \boldsymbol{h}|\boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} e^{-\Phi(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta})}$

$P(\boldsymbol{v}_n|\boldsymbol{\theta})$ をどう定義するか？
$\rightarrow$ $\boldsymbol{h}$ を周辺化する

$$P(\boldsymbol{v}_n|\boldsymbol{\theta}) = \sum_{\boldsymbol{h}} P(\boldsymbol{v}_n, \boldsymbol{h}|\boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} \sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}$$

対数尤度関数：$\ln L(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln P(\boldsymbol{v}_n|\boldsymbol{\theta})$

$$= \sum_{n=1}^{N} \ln\left(\sum_{\boldsymbol{h}} P(\boldsymbol{v}_n, \boldsymbol{h}|\boldsymbol{\theta})\right)$$

$$= \sum_{n=1}^{N} \left(\ln\left(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right) - \ln z(\boldsymbol{\theta})\right)$$

11/16/2023

12

対数尤度関数の偏微分

- $\frac{\partial \ln L(\boldsymbol{\theta})}{\partial w_{j,i}} = \sum_{n=1}^{N} \frac{\partial\left(\ln\left(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right) - \ln z(\boldsymbol{\theta})\right)}{\partial w_{j,i}}$

$$= \sum_{n=1}^{N} \left(\boxed{\frac{\partial\left(\ln \sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right)}{\partial w_{j,i}}} - \boxed{\frac{\partial \ln z(\boldsymbol{\theta}))}{\partial w_{j,i}}} \right) \quad E_{\boldsymbol{\theta}}[x_i x_j]$$

$$\boxed{\frac{\partial\left(\ln \sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right)}{\partial w_{j,i}}} = \frac{1}{\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}} \sum_{\boldsymbol{h}} \frac{\partial e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}}{\partial w_{j,i}}$$

$$= \frac{1}{P(\boldsymbol{v}_n|\boldsymbol{\theta}) z(\boldsymbol{\theta})} \sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})} \frac{\partial(-\Phi(\boldsymbol{v}_n, \boldsymbol{h}, \boldsymbol{\theta}))}{\partial w_{j,i}}$$

$$= \sum_{\boldsymbol{h}} \frac{P(\boldsymbol{v}_n, \boldsymbol{h}|\boldsymbol{\theta}) z(\boldsymbol{\theta})}{P(\boldsymbol{v}_n|\boldsymbol{\theta}) z(\boldsymbol{\theta})} \boxed{x_i^n x_j^n} = \sum_{\boldsymbol{h}} \frac{P(\boldsymbol{v}_n, \boldsymbol{h}|\boldsymbol{\theta})}{P(\boldsymbol{v}_n|\boldsymbol{\theta})} x_i^n x_j^n$$

可視ユニットなら $\boldsymbol{v}_n$ のユニット値,
隠れユニットならその時の $\boldsymbol{h}$ のユニット値

$$= \sum_{\boldsymbol{h}} P(\boldsymbol{h}|\boldsymbol{v}_n, \boldsymbol{\theta}) x_i^n x_j^n$$

11/16/2023

13

対数尤度関数の偏微分

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial w_{j,i}} = \sum_{n=1}^{N} \frac{\partial\left(\ln\left(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right) - \ln z(\boldsymbol{\theta})\right)}{\partial w_{j,i}}$$

$$= \sum_{n=1}^{N}\left(\boxed{\frac{\partial\left(\ln \sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}\right)}{\partial w_{j,i}}} - \boxed{\frac{\partial \ln z(\boldsymbol{\theta})}{\partial w_{j,i}}}\right) \quad E_{\boldsymbol{\theta}}[x_i x_j]$$

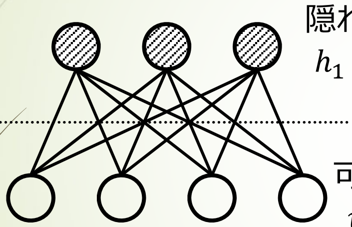
$$= \sum_{n=1}^{N}\left(\sum_{\boldsymbol{h}} P(\boldsymbol{h}|\boldsymbol{v}_n,\boldsymbol{\theta}) x_i^n x_j^n - E_{\boldsymbol{\theta}}[x_i x_j]\right)$$

$$= \sum_{n=1}^{N}\left(\sum_{\boldsymbol{h}} P(\boldsymbol{h}|\boldsymbol{v}_n,\boldsymbol{\theta}) x_i^n x_j^n\right) - N E_{\boldsymbol{\theta}}[x_i x_j]$$

可視ユニットなら x_i^n は $\boldsymbol{v}_n$ のユニット値，
隠れユニットならその時の $\boldsymbol{h}$ のユニット値
（x_i^n，x_j^nが全て可視ユニットの場合，
通常のボルツマンマシンと同じになる）

14

制限ボルツマンマシン（Restricted Boltzmann Machine: RBM）



隠れユニット
$h_1 \cdots h_{|H|} \in H$

可視ユニット
$v_1 \cdots v_{|V|} \in V$

※可視ユニット同士，隠れユニット同士の結合のないボルツマンマシン

エネルギー関数 $\Phi(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta}) = -\sum_{(i,j)} w_{j,i} v_i h_j - \sum_i a_i v_i - \sum_j b_j h_j$

バイアスを二つに分けた

15

RBMの対数尤度関数

- $\ln L(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln P(\boldsymbol{v}_n|\boldsymbol{\theta})$
  $= \sum_{n=1}^{N} \left(\boxed{\ln(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n, \boldsymbol{h}, \boldsymbol{\theta})})} - \ln z(\boldsymbol{\theta}) \right)$

$\boxed{\ln(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n, \boldsymbol{h}, \boldsymbol{\theta})})}$

$= \ln(\sum_{\boldsymbol{h}} e^{\sum_{(i,j)} w_{j,i} v_i^n h_j + \sum_i a_i v_i^n + \sum_j b_j h_j})$

$= \ln(\sum_{\boldsymbol{h}} e^{\sum_{(i,j)} w_{j,i} v_i^n h_j} e^{\sum_i a_i v_i^n} e^{\sum_j b_j h_j})$

$= \ln(e^{\sum_i a_i v_i^n} \sum_{\boldsymbol{h}} e^{\sum_{(i,j)} w_{j,i} v_i^n h_j} e^{\sum_j b_j h_j})$

$= \ln(e^{\sum_i a_i v_i^n}) + \ln(\sum_{\boldsymbol{h}} e^{\sum_{(i,j)} w_{j,i} v_i^n h_j} e^{\sum_j b_j h_j})$

$= \sum_i a_i v_i^n + \ln(\sum_{\boldsymbol{h}} e^{\sum_j h_j (b_j + \sum_i w_{j,i} v_i^n)})$

16

RBMの対数尤度関数

- $\ln(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n, \boldsymbol{h}, \boldsymbol{\theta})})$
  $= \sum_i a_i v_i^n + \ln(\boxed{\sum_{\boldsymbol{h}} e^{\sum_j h_j (b_j + \sum_i w_{j,i} v_i^n)}})$

$\boxed{\sum_{\boldsymbol{h}} e^{\sum_j h_j (b_j + \sum_i w_{j,i} v_i^n)}}$

$= \sum_{\boldsymbol{h}} e^{h_1(b_1 + \sum_i w_{1,i} v_i^n) + h_2(b_2 + \sum_i w_{2,i} v_i^n) + \cdots}$

$= \sum_{\boldsymbol{h}} e^{h_1(b_1 + \sum_i w_{1,i} v_i^n)} e^{h_2(b_2 + \sum_i w_{2,i} v_i^n)} \ldots$

$= \sum_{h_1=0,1} \sum_{h_2=0,1} \cdots \sum_{h_{|H|}=0,1} e^{h_1(b_1 + \sum_i w_{1,i} v_i^n)} \ldots$

$= \sum_{h_1=0,1} e^{h_1(b_1 + \sum_i w_{1,i} v_i^n)} \sum_{h_2=0,1} e^{h_2(b_2 + \sum_i w_{2,i} v_i^n)} \ldots$

$= (1 + e^{b_1 + \sum_i w_{1,i} v_i^n})(1 + e^{b_2 + \sum_i w_{2,i} v_i^n}) \ldots$

$= \prod_j (1 + e^{b_j + \sum_i w_{j,i} v_i^n})$

※ユニット値が {0,1} でない場合は異なる導出結果になる

17

RBMの対数尤度関数

- $\ln(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})})$
  $= \sum_i a_i v_i^n + \ln(\boxed{\sum_{\boldsymbol{h}} e^{\sum_j h_j(b_j+\sum_i w_{j,i} v_i^n)}})$
  $= \sum_i a_i v_i^n + \ln(\prod_j (1+e^{b_j+\sum_i w_{j,i} v_i^n}))$
  $= \sum_i a_i v_i^n + \sum_j \ln(1+e^{b_j+\sum_i w_{j,i} v_i^n})$
- $\ln L(\boldsymbol{\theta}) = \sum_{n=1}^{N}(\ln(\sum_{\boldsymbol{h}} e^{-\Phi(\boldsymbol{v}_n,\boldsymbol{h},\boldsymbol{\theta})}) - \ln z(\boldsymbol{\theta}))$
  $= \sum_{n=1}^{N}\left(\sum_i a_i v_i^n + \sum_j \ln(1+e^{b_j+\sum_i w_{j,i} v_i^n}) - \ln z(\boldsymbol{\theta})\right)$

11/16/2023

18

RBMの対数尤度関数の偏微分

- $\ln L(\boldsymbol{\theta})$
  $= \sum_{n=1}^{N}\left(\sum_i a_i v_i^n + \sum_j \ln(1+e^{b_j+\sum_i w_{j,i} v_i^n}) - \ln z(\boldsymbol{\theta})\right)$
- $\frac{\partial \ln L(\boldsymbol{\theta})}{\partial w_{j,i}}$
  $= \sum_{n=1}^{N}\left(\underbrace{\boxed{\frac{\partial \sum_i a_i v_i^n}{\partial w_{j,i}}}}_{0} + \boxed{\frac{\partial \sum_j \ln(1+e^{b_j+\sum_i w_{j,i} v_i^n})}{\partial w_{j,i}}} - \underbrace{\boxed{\frac{\partial \ln z(\boldsymbol{\theta})}{\partial w_{j,i}}}}_{E_{\boldsymbol{\theta}}[v_i h_j]}\right)$

$$\boxed{\frac{\partial \sum_j \ln(1+e^{b_j+\sum_i w_{j,i} v_i^n})}{\partial w_{j,i}}} = \frac{1}{1+e^{b_j+\sum_i w_{j,i} v_i^n}} e^{b_j+\sum_i w_{j,i} v_i^n} \frac{\partial (b_j+\sum_i w_{j,i} v_i^n)}{\partial w_{j,i}}$$

11/16/2023

19

RBMの対数尤度関数の偏微分

$$\boxed{\frac{\partial \sum_j \ln\left(1+e^{b_j+\sum_i w_{j,i} v_i^n}\right)}{\partial w_{j,i}}}$$

$$= \frac{1}{1+e^{b_j+\sum_i w_{j,i} v_i^n}} e^{b_j+\sum_i w_{j,i} v_i^n} \frac{\partial\left(b_j+\sum_i w_{j,i} v_i^n\right)}{\partial w_{j,i}}$$

$$= \frac{e^{b_j+\sum_i w_{j,i} v_i^n}}{1+e^{b_j+\sum_i w_{j,i} v_i^n}} v_i^n$$

ここで $\lambda_j^n = b_j + \sum_i w_{j,i} v_i^n$ とおくと

$$= \frac{e^{\lambda_j^n}}{1+e^{\lambda_j^n}} v_i^n = \frac{1}{1+e^{-\lambda_j^n}} v_i^n$$

$\left(P\left(h_j = 1\right) = \frac{1}{1+e^{-\lambda_j}}\right)$

⇧

隠れ層のユニット同士の結合はないため，$h_j = 1$ の確率は v_i の値のみに依存して決まる

20

RBMの対数尤度関数の偏微分

- $\frac{\partial \ln L(\boldsymbol{\theta})}{\partial w_{j,i}}$

$$= \sum_{n=1}^{N} \left(\overset{0}{\boxed{\frac{\partial \sum_i a_i v_i^n}{\partial w_{j,i}}}} + \boxed{\frac{\partial \sum_j \ln\left(1+e^{b_j+\sum_i w_{j,i} v_i^n}\right)}{\partial w_{j,i}}} - \overset{E_{\boldsymbol{\theta}}[v_i h_j]}{\boxed{\frac{\partial \ln z(\boldsymbol{\theta})}{\partial w_{j,i}}}} \right)$$

$$= \sum_{n=1}^{N} \left(\frac{1}{1+e^{-\lambda_j^n}} v_i^n - E_{\boldsymbol{\theta}}[v_i h_j] \right)$$

$$= \sum_{n=1}^{N} \left(\frac{1}{1+e^{-\lambda_j^n}} v_i^n \right) - N E_{\boldsymbol{\theta}}[v_i h_j]$$

出題予定の演習課題

- 制限ボルツマンマシンの対数尤度関数の偏微分