機械学習（8回目）

創域理工学部　情報計算科学科
桂田　浩一

前回の復習

- ホップフィールドモデル

3

本日の内容

- ボルツマンマシン

4

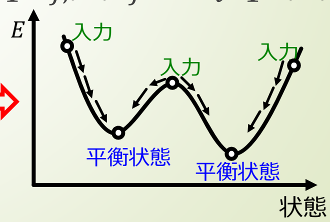
ホップフィールドモデルの問題点

- ホップフィールドモデルのユニット値更新式

$$x_i(t+1) = \begin{cases} 1 & (u_i(t) > 0) \\ x_i(t) & (u_i(t) = 0) \\ 0 & (u_i(t) < 0) \end{cases} \qquad u_i(t) = \sum_{j=1}^{n} w_{j,i} x_j(t) - \theta_i$$

- x_i を更新すると $E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{j,i}x_ix_j + \sum_{i=1}^{n}\theta_ix_i$ が小さくなっていく

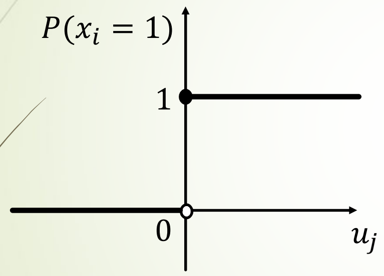
極小点に収束して最小点に行かない ⇨



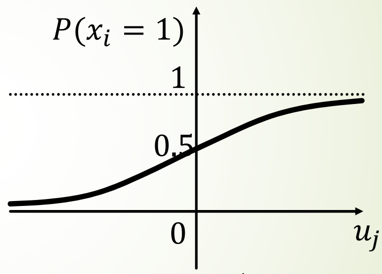
5

ランダム性の導入

- $x_i = 1$ になる確率を $P(x_i = 1)$ と表したとき



ホップフィールドモデル



$P(x_i = 1) = \dfrac{1}{1 + e^{-u_i/T}}$ とする

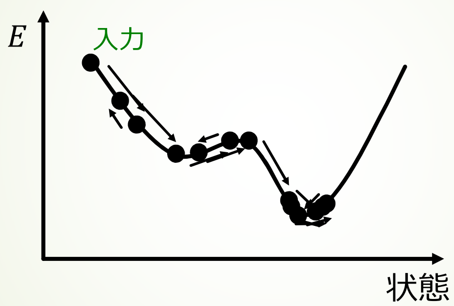
$u_i > 0$でも必ず$x_i(t+1) = 1$とは限らない

10/19/2023

6

エネルギー関数の動き

- x_i を更新していったときの E の動き



※一つの値には収束しない
※最小点を中心に確率的な状態の分布になる

10/19/2023

7

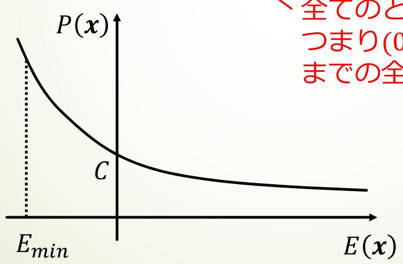
最終的なx_iの分布

- 状態を $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$ とするとき

$$P(\boldsymbol{x}) = Ce^{-E(\boldsymbol{x})/T}$$ （ボルツマン分布）

ただし $C = \frac{1}{\sum_{\boldsymbol{x}} e^{-E(\boldsymbol{x})/T}} \left(\cong \frac{1}{\int e^{-E(\boldsymbol{x})/T} dE(\boldsymbol{x})}\right)$

全てのとりうる（$x_1, x_2, \cdots, x_m$），
つまり$(0,0,\cdots,0)$から$(1,1,\cdots,1)$
までの全てのパターン



10/19/2023

8

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

10/19/2023

9

今後の説明のための記号の変更

- $E = -\frac{1}{2}\sum_i \sum_j w_{j,i} x_i x_j + \sum_i \theta_i x_i$
  $= -\sum_{(i,j)} w_{j,i} x_i x_j + \sum_i \theta_i x_i$ 　　（(i,j) は異なるi,jの組合せ）
- $P(\boldsymbol{x}) = Ce^{-E(\boldsymbol{x})/T}$ 　　$\left(C = \frac{1}{\sum_x e^{-E(\boldsymbol{x})/T}}\right)$

バイアス
⇩
ここで $T = 1$，$b_i = -\theta_i$，$\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{b})$，$z(\boldsymbol{\theta}) = \frac{1}{C}$ と置き，
$\Phi(\boldsymbol{x}, \boldsymbol{\theta}) = E$ とすると

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

$\boldsymbol{\theta}$ をパラメータと考え，$\boldsymbol{\theta}$ が与えられたときに
$\boldsymbol{x}$ を取る確率を考える

10/19/2023

10

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！！

10/19/2023

11

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

10/19/2023

12

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： $b \leftarrow b + \eta \frac{\partial \ln L(\boldsymbol{\theta})}{\partial b}$

10/19/2023

13

対数尤度関数の偏微分

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

- $\boxed{-\frac{\partial \Phi(\boldsymbol{x}_n,\boldsymbol{\theta})}{\partial w_{j,i}}} = \frac{\partial\left(\sum_{(i,j)} w_{j,i}x_i^n x_j^n + \sum_i b_i x_i^n\right)}{\partial w_{j,i}} = x_i^n x_j^n$

14

対数尤度関数の偏微分

- $\boxed{-\frac{\partial \Phi(\boldsymbol{x}_n,\boldsymbol{\theta})}{\partial w_{j,i}}} = \frac{\partial\left(\sum_{(i,j)} w_{j,i}x_i^n x_j^n + \sum_i b_i x_i^n\right)}{\partial w_{j,i}} = x_i^n x_j^n$

- $\boxed{-\frac{\partial \ln z(\boldsymbol{\theta})}{\partial w_{j,i}}} = -\frac{1}{z(\boldsymbol{\theta})}\frac{\partial z(\boldsymbol{\theta})}{\partial w_{j,i}} = -\frac{1}{z(\boldsymbol{\theta})}\frac{\partial \sum_{\boldsymbol{x}} e^{-\Phi(\boldsymbol{x},\boldsymbol{\theta})}}{\partial w_{j,i}}$

$$= -\frac{1}{z(\boldsymbol{\theta})}\sum_{\boldsymbol{x}}\frac{\partial e^{-\Phi(\boldsymbol{x},\boldsymbol{\theta})}}{\partial w_{j,i}} = -\frac{1}{z(\boldsymbol{\theta})}\sum_{\boldsymbol{x}} e^{-\Phi(\boldsymbol{x},\boldsymbol{\theta})}\frac{\partial(-\Phi(\boldsymbol{x},\boldsymbol{\theta}))}{\partial w_{j,i}}$$

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

全ての $\boldsymbol{x}$ を考えたときの $x_i x_j$ の期待値
（$E_{\boldsymbol{\theta}}[x_i x_j]$と表す）

15

対数尤度関数の偏微分

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

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

$$= \sum_{n=1}^{N}(x_i^n x_j^n - E_{\boldsymbol{\theta}}[x_i x_j])$$
$$= \sum_{n=1}^{N}(x_i^n x_j^n) - N E_{\boldsymbol{\theta}}[x_i x_j]$$

10/19/2023

16

期待値の効率的な求め方

- ギブスサンプリング

$$E_{\boldsymbol{\theta}}[x_i x_j] = \sum_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{\theta}) x_i x_j$$

全ての $\boldsymbol{x}$ について加算するのは困難
⇒ $\boldsymbol{x}$ をランダムに $P(\boldsymbol{x}|\boldsymbol{\theta})$ の確率で生成して平均値を求める．
（$P(x_i = 1) = \frac{1}{1+e^{-u_i}}$ を使う．）

メリット：少ない数の x の生成で期待値の近似値が求まる

10/19/2023

出題予定の演習課題

- 対数尤度関数の偏微分