Subject 18H_Q2-7 Date 22

Attendants

$x \in \mathbb{R}^D$ : observation

$z \in \mathbb{R}^K$ : latent variable

$$p(x,z) = \frac{1}{Z} \exp\left(-E(x,z)\right)$$

正規化定数 ← $Z$

energy function ← $E(x,z)$

parameterized by DNN

training dataset of $N$ example ( $x^{(i)}$ with $i \in [1,N]$ )

gradient descent の学習過程を記述せよ

( log probability of $p(x)$ 対数尤度を最大化するように )

Answer this problem can be formed as follows

$$\hat{\theta} = \arg\max_{\theta} p_\theta(x,z)$$

$$= \arg\max_{\theta} \log p(x,z)$$

by using this grad. run GD.

$$\frac{\partial \log(p_\theta(x,z))}{\partial \theta} = \mathbb{E}_{p_{data}}\left[\frac{\partial E}{\partial \theta}\right] - \mathbb{E}_{p_\theta(x)}\left[\frac{\partial E}{\partial \theta}\right]$$

Subject EBM : Energy-Based Model Date 23

Attendants

xを観測 → 変数の集合 yを出力

EBMは xと相性の良い yを見つける問題を考える

$$\hat{y} = \underset{y}{\operatorname{argmin}} F(x, y)$$

$F : x \times y \to \mathbb{R}$

F(x,y)が低い yを見つける問題

e.g. テキスト yは テキスト xの良い翻訳であるか？

同定法による推論 、 互換性のある yを見つける

---

潜在変数を持つEBM z = latent variable

$$y, z = \underset{y, z}{\operatorname{argmin}} F(x, y, z)$$

Subject Date 24

Attendants

データ $x$ の確率密度関数 $p_\theta(x)$

$$p_\theta(x) = \frac{\exp(-E_\theta(x))}{Z(\theta)}$$

where $Z(\theta) = \int \exp(-E_\theta(x))\, dx$

$Z$ 分配関数

$E$ : エネルギー関数

密度比推定

エネルギー関数は密度の対数の負の値に比例

ただし $Z(\theta)$ があるため密度自体は計算できない

↳ NCE で推定はできる

どう学習するか？

→ 最尤推定 train data に対する対数尤度 $\log p_\theta(x)$ を最大化する

↳ $Z(\theta)$ の計算を避けたい

Subject: Contrastive Divergence Learning　Date: 2月

Attendants: CD法

SGDでの学習の場合，対数尤度そのものではなく

パラメータ$\theta$での勾配が必要なのだ！

$$\frac{\partial \log p_\theta(x)}{\partial \theta}$$

$$p_\theta = \frac{\exp(-E_\theta(x))}{\int \exp(-E_\theta(x))dx}$$

$\log(x \div y) = \log x - \log y$

$$= -\mathbb{E}_{p_\theta(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right] + \mathbb{E}_{p_{data}(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right]$$

$$\log p_\theta(x) = \log_e e^{-E_\theta(x)} - \log \int e^{-E_\theta(x)} dx$$

$$= -E_\theta(x) + \int E_\theta(x) dx$$

$$\frac{\partial \log p_\theta(x)}{\partial \theta} = \mathbb{E}_{p_{data}(x)}\left[\frac{\partial E_\theta(x)}{\partial x}\right]$$ データ分布上の勾配の期待値

$$- \mathbb{E}_{p_\theta(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right]$$ モデル分布上の勾配の期待値

$p_\theta(x)$からのサンプルはMCMCとかでとってくる？

↳ 大変なので $p_\theta(x)$からサンプルをとりたくない

Subject NCE : Noise Contrastive Estimation Date 2-6

Attendants

# ノイズ対照推定

分配関数 $Z(\theta)$ 自体を別のパラメータ $c$ で推定

$\log p_\theta(x) = -E_\theta(x) - c$ を最大化する

$$J(\theta) = \mathbb{E}_{p_{data}(x)}\left[\log \frac{p_\theta(x)}{p_\theta(x)+q(x)}\right]$$

$$+ \mathbb{E}_{q(x)}\left[\log \frac{q(x)}{p_\theta(x)+q(x)}\right]$$

$q(x)$ はなんかしらの
ノイズ分布、e.g
Gaussian,

$J(\theta)$ の最大化が $\theta$ は対数尤度が最大化され
$c = Z(\theta)$ に一致するらしい。

直感的には データからのサンプルとノイズを
見分けるように学習
実はGANと近い関係

Subject Date 27

Attendants

EBMは 計算神経科学での脳のモデルとして
いまだによく使われる

最近のEBM：エネルギー関数
そのものをNNで定義
↳全体での1つの決定的な関数

エネルギー関数の出力をそのまま用いて学習