Deep Learning 輪読会 2017

第17章　モンテカルロ法

2018.01.15

理学系研究科附属
天文学教育研究センター
学部4年 吉村勇紀

構成

17.1 サンプリングとモンテカルロ法

17.2 重点サンプリング

17.3 MCMC

17.4 ギブスサンプリング

17.5 モードの混合問題

17.1 サンプリングとモンテカルロ法

- モンテカルロサンプリング
  - 総和や積分をある確率分布の下での期待値に置き換える

$$s = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) f(\boldsymbol{x}) = E_p[f(\mathbf{x})]$$

$$\hat{s}_n = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}^{(i)})$$

  - これは不偏推定量である

$$\mathbb{E}[\hat{s}_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[f(\boldsymbol{x}^{(i)})] = \frac{1}{n} \sum_{i=1}^{n} s = s$$

  - 分散はn無限大で0に

$$\begin{aligned} \mathrm{Var}[\hat{s}_n] &= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[f(\mathbf{x})] \\ &= \frac{\mathrm{Var}[f(\mathbf{x})]}{n}. \end{aligned}$$

17.2 重点サンプリング

- 被積分（被総和）関数の分解
  - 一通りではない
  - モンテカルロ推定量は重点サンプリング推定量に変換できる

$$\hat{s}_q = \frac{1}{n} \sum_{i=1, \mathbf{x}^{(i)} \sim q}^{n} \frac{p(\boldsymbol{x}^{(i)}) f(\boldsymbol{x}^{(i)})}{q(\boldsymbol{x}^{(i)})}$$

  - 両推定量の期待値はqによらず一致 $\mathbb{E}_q[\hat{s}_q] = \mathbb{E}_q[\hat{s}_p] = s$
  - 推定量の分散はqの選択に敏感で次のとき最小

$$q^*(\boldsymbol{x}) = \frac{p(\boldsymbol{x})|f(\boldsymbol{x})|}{Z}$$

- 無バイアス重点サンプリング
  - 正規化を必要としない

17.3 マルコフ連鎖モンテカルロ法

- MCMC
  - マルコフ連鎖を用いたサンプリング法
  - マルコフ連鎖 -> 条件付き独立な確率に従う系列
  - 遷移確率 T(x'|x) による確率的状態遷移
  - 条件はどの状態にたいしても確率0ではない -> EBM
  - EBMから同時分布をサンプリングするのは容易でない -> 卵が先か鶏が先か
- 平衡状態への収束
  - 遷移は行列、確率はベクトルで書ける $\boldsymbol{v}^{(t)} = \boldsymbol{A}\boldsymbol{v}^{(t-1)}$
  - 遷移行列の固有値のうち1未満のものは0に収束
  - $\boldsymbol{v}^{(t)} = \left(\boldsymbol{V}\mathrm{diag}(\boldsymbol{\lambda})\boldsymbol{V}^{-1}\right)^{t}\boldsymbol{v}^{(0)} = \boldsymbol{V}\mathrm{diag}(\boldsymbol{\lambda})^{t}\boldsymbol{V}^{-1}\boldsymbol{v}^{(0)}$

17.3 マルコフ連鎖モンテカルロ法

- MCMCのコスト
  - 平衡状態に達する（burning in）までの時間（混合時間）
  - 平衡状態から独立なサンプルを得るのにかかるステップ数

- MCMCの困難
  - 混合時間が不明である
  - 定常状態に達したか厳密にテストできない
  - ヒューリスティックな方法でやるしかない

17.4 ギブスサンプリング

- 遷移確率の決め方
  - 学習したモデル分布から導く -> ギブスサンプリング
  - 遷移確率を直接学習する -> 20章
- ギブスサンプリング

1. Initialize $\{z_i : i = 1, \ldots, M\}$
2. For $\tau = 1, \ldots, T$:
   - Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.
   - Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$.

     $\vdots$
   - Sample $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$.

  - メトロポリス・ヘイスティング法の一種
    - 以下の確率で新たなサンプルを採択する棄却サンプリング

$$A_k(\mathbf{z}^\star, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^\star) q_k(\mathbf{z}^{(\tau)} | \mathbf{z}^\star)}{\widetilde{p}(\mathbf{z}^{(\tau)}) q_k(\mathbf{z}^\star | \mathbf{z}^{(\tau)})}\right)$$

17.5 分離されたモード間の混合の課題

- 不十分な混合
  - エネルギーの低い状態への遷移 -> 小ステップで実現
  - エネルギーの高い状態への遷移 -> ランダムウォーク的（非効率）
  - 一旦あるモードに達すると他のモードへ移動しにくい

17.5 分離されたモード間の混合の課題

- 焼きなまし法
  - 目標分布を変える
  - EBMの場合逆温度βを導入する

$$p_\beta(\boldsymbol{x}) \propto \exp\left(-\beta E(\boldsymbol{x})\right)$$

  - β<1 (高温) でサンプリングして混合した後、β=1で混合する
  - 様々な温度で同時に行う並列焼きなまし法もある

$$p(T,E,T',E') = \begin{cases} 1 & if \quad \Delta T \cdot \Delta E < 0 \\ \exp\left(-\frac{\Delta T \cdot \Delta E}{T \cdot T'}\right) & otherwise \end{cases}$$

17.5 分離されたモード間の混合の課題

- 深さが混合に役立つ可能性
  - $p(h|x)$においてhの表現を深くする
  - 深い自己符号下記やRBMでは異なるモードの隔たりが小さい
  - 深層モデルによるサンプリング法の支援については不明なことが多い

参考文献

- Deep Learning
  - Ian Goodfellow, Yoshua Bengio, Aaron Courville
  - 日本語版 https://www.amazon.co.jp/%E6%B7%B1%E5%B1%A4%E5%AD%A6%E7%BF%92-Ian-Goodfellow/dp/4048930621