

leaf node

↓

back propagation

RNN w/ storing activation

↓

forward node is

the one

to allow you

to calc gradients

How does it work?

↓
unroll then (simple way).

What is a draw back?

↑
缺点的

$$\text{minimize } J(\theta) = \mathbb{E}_{X \sim p} [h(X; \theta)]$$

where X is drawn from a dist p which has no dependence on $\theta \in \mathbb{R}^n$

$$\mathbb{E}_{X \sim p} [h(X; \theta)] \simeq \frac{1}{m} \sum_{i=1}^m h(X_i; \theta)$$

$$X_i \sim p \text{ i.i.d}$$

where m is number of samples

Answer

(a) If p is known, but assuming that you can draw samples from it, explain how could approximate the above expectation.

(b) Assuming that h is differentiable in θ , write an algorithm to minimize J from samples.

$$\text{Full Batch Gradient } \theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta)$$

$$\text{where } J(\theta) = \frac{1}{m} \sum_{i=1}^m \nabla_{\theta} h(X_i; \theta), \eta \text{ is step size.}$$

1. (c) How do we call the approach developed in this question within the broader context of ML?

Map out the term from the general problem formulation above to the specific problem setting of ML. Define

[i] The random variables over which the expectation is taken

X is called input data or feature

[ii] h is called loss function

J is called generalization loss or true loss

[iii] The optimization variables θ is called model parameter.

(d) Some dist q is available instead of p and assume that you can evaluate p and q everywhere, then
How do you estimate the above expectation

$$\mathbb{E}_{x \sim p} [h(x; \theta)]$$

$\text{supp}(f) := \{x \in X \mid f(x) \neq 0\}$
Supported in $(\frac{1}{2}, \frac{3}{2})$

$$= \int_{x \in \text{supp}(p)} p(x) h(x; \theta) dx$$

$$= \int_{x \in \text{supp}} q(x) \frac{p(x)}{q(x)} h(x; \theta) dx$$

$$= \mathbb{E}_{x \sim q} \left[\frac{p(x)}{q(x)} h(x; \theta) \right] \approx \frac{1}{m} \sum_{i=1}^m \frac{q(x_i)}{p(x_i)} h(x_i; \theta)$$

as condition

$$q(x) > 0 \quad \forall x \in \text{supp}(p)$$

$$x_i \sim q \text{ i.i.d}$$

1. Monte Carlo

(e) Can you think a concrete problem setting in which the situation described in §1.1d1

Answer,

The Out-of-Distribution setting is one of concrete problem.

we aim to generalize model on unseen distribution p

by using data which is drawn from different distribution q

↑

この解答では p と q が異なる分布であることを示す

使用できる...

→ 累積分布関数 → 重点- γ - γ^0 - γ^1

Importance Sampling

$p(x)$ は pdf である

ある量 $f(x)$ の平均値

$$I = \int f(x) p(x) dx \text{ に対する問題}$$

モンテカルロ積分 (Monte Carlo Integration)

$p(x)$ から N 個のサンプル $x_1, \dots, x_N$ を生成する

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad \text{に対する問題}$$

Importance Sampling

$$I = \int f(x) p(x) dx = \int f(x) \frac{p(x)}{q(x)} q(x) dx$$

$$\therefore \text{ここで } f'(x) = f(x) \frac{p(x)}{q(x)} \text{ である}$$

$$I = \int f'(x) q(x) dx \approx \frac{1}{N} \sum_{i=1}^N f'(x_i) \quad x_i \sim q$$

$$p(y) = p(z) \left| \frac{dz}{dy} \right|$$

$y = f(z)$ みたいな定式化がしたい..
 $\nearrow$ f について必要なのは、

$z \sim U(0,1)$ ことの分布は

$y \sim f(z)$

$$p(y) = p(z) \left| \frac{dz}{dy} \right|$$

$\uparrow$

変数変換の
ために z が出て

e.g. $\int p(y) dy$



みたいな z が必要になる

今 $p(y)$ は $f(z)$ で表したい.. z で積分したい z -

$\frac{dz}{dy}$ を求む (今は strict に f は z だけ) $(f$ は strict に z だけ)

$$p(y) = p(z) \left| \frac{dz}{dy} \right|$$

$$p_z(y) = p_z(z) \left| \frac{dz}{dy} \right| \text{ と } dz = c(dy)$$

$z \sim U(z|0,1)$ 仮定

$$p_z(z) = 1$$



z は 100% 一様

これは仮定

PDF と U の表

5-2

$$p_z(y) = \left| \frac{dz}{dy} \right|$$

$$z = h(y) \text{ と}$$

$$0 < z$$

$$p_z(y) = \left| \frac{d h(y)}{dy} \right| = |h'(y)|$$

z は $h(y)$ の PDF $p_z(z)$ と変換

$$h(y) = \int_{-\infty}^y p_z(t) dt$$

z の $h(t)$ は Cumulative Dist Func の z の

逆関数

$$P_y(y) = |h'(y)| = \left| \frac{d}{dy} \int_{-\infty}^y P_x(t) dt \right|$$

$$= |P_x(x)| = P_x(x)$$

以上より $P_y(y) = P_x(x)$ となる

$z = h(y) = \int_{-\infty}^y P_y(t) dt$ は y の Cumulative Dist Func である

$$y = h^{-1}(z) \quad z$$

z が P_y に従って変化する y の変換である

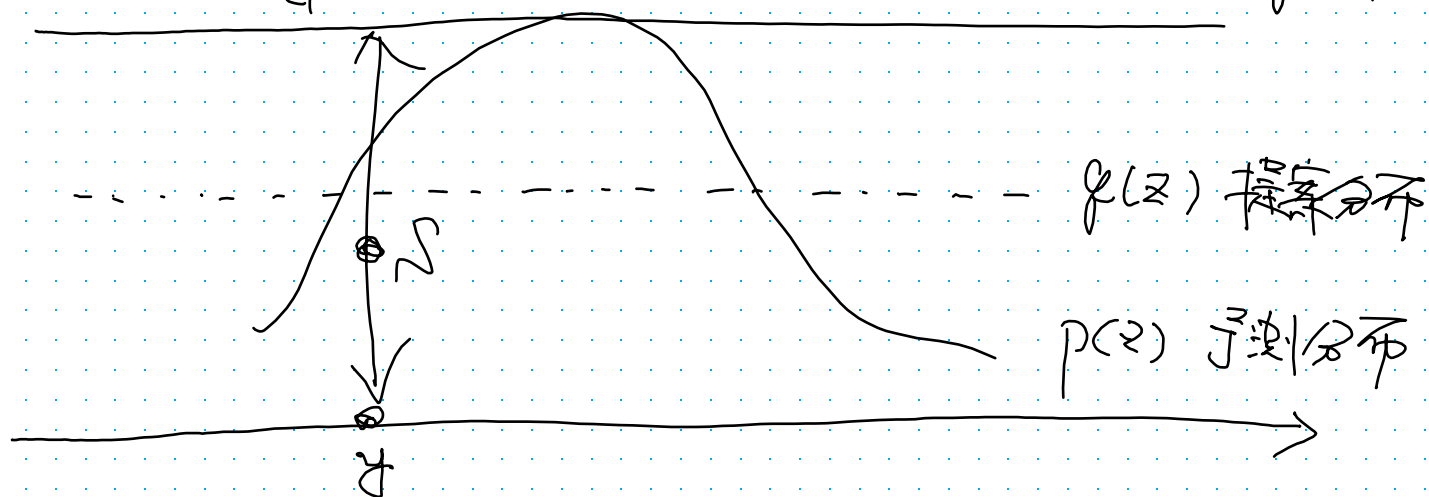
~~棄却~~ $p(z) \propto \pi(z)$

Motivation $p(z)$ からの $\pi(z)$ は難しいから

Bayes' rule

$Mq(z)$

$p(z) = \frac{1}{Z} \tilde{p}(z)$ の $\tilde{p}(z)$ から容易に



1. 提案分布 $q(z)$ に従う乱数 y を発生

↳ 上の図では、 z は $q(z)$ に従う乱数

2. $[0, Mq(y)]$ に従う乱数 S を発生

(この M は $p(z)/q(z)$ の最大値)

3. $S < p(z)$ ならば受入れ

otherwise 棄却

4. loop 1~3

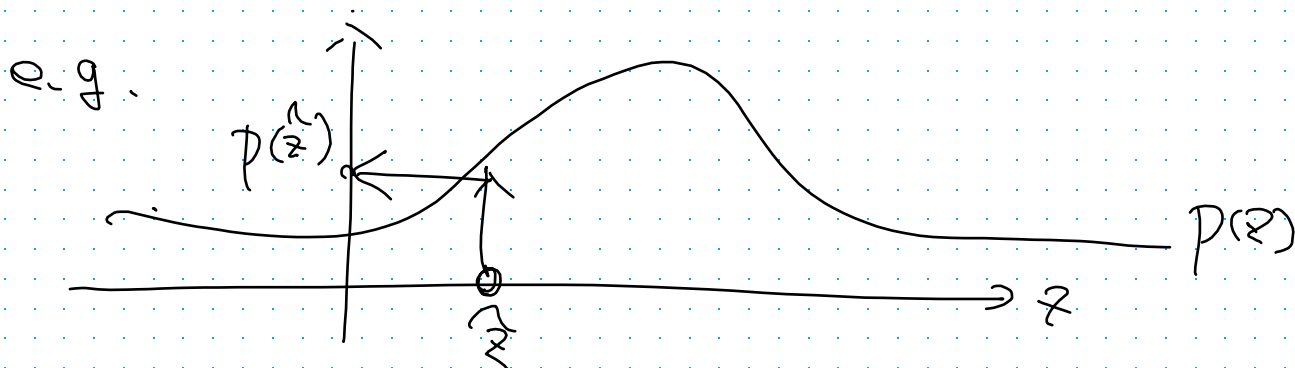
重点サンプリング (Importance Sampling)

期待値を直接近似する方法

Assumption: $p(z)$ の複雑 (サンプリ) 回避

任意の与えられた z に対して $p(z)$

計算 $\theta = \epsilon_1 z^2 \pm z!$



上 z に対して $p(z)$ を求めたい

この場合 z の Samp (サンプリング) は z の分布を $q(z)$ とする

z を $q(z)$ から $p(z)$ の分布に変換する

$$\begin{aligned} \mathbb{E}_{z \sim p(z)} [f] &= \int p(z) f(z) dz && \text{importance weights} \\ &= \int q(z) \frac{p(z)}{q(z)} f(z) dz = \mathbb{E}_{z \sim q(z)} \left[f(z) \frac{p(z)}{q(z)} \right] \end{aligned}$$

Importance Sampling

$$\mathbb{E}_{z \sim p(z)} [f] \approx \mathbb{E}_{z \sim q(z)} \left[\frac{p(z)}{q(z)} f(z) \right]$$

$$= \frac{1}{L} \sum_{l=1}^L \frac{p(z_l)}{q(z_l)} f(z_l) \quad \dots *$$

$$p(z) = \frac{\tilde{p}(z)}{Z_p}$$

$$q(z) = \frac{\tilde{q}(z)}{Z_q}$$

$$Z_p = \int \tilde{p}(z) dz$$

$$* = \frac{1}{L} \frac{Z_q}{Z_p} \sum_{l=1}^L \frac{\tilde{p}(z_l)}{\tilde{q}(z_l)} f(z_l)$$

$$\frac{1}{Z_q} = \frac{q(z)}{\tilde{q}(z)}$$

2つの式を掛け合わせると

$$\frac{Z_p}{Z_q} = \frac{1}{Z_q} \int \tilde{p}(z) dz = \int \frac{\tilde{p}(z)}{\tilde{q}(z)} q(z) dz$$

$$\approx \frac{1}{L} \sum_{l=1}^L \frac{\tilde{p}(z_l)}{\tilde{q}(z_l)} = Q$$

困難な場合

$$L \neq M, z$$

$$\mathbb{E}_{z \sim p(z)} \approx \frac{z_q}{z_p} \frac{1}{L} \sum_{l=1}^L \frac{\tilde{p}(z_l)}{\tilde{q}(z_l)} f(z_l)$$

$$= \frac{1}{Q} \frac{1}{L} \sum_{l=1}^L \frac{\tilde{p}(z_l)}{\tilde{q}(z_l)} f(z_l)$$

$$\mathbb{E}_{z \sim p(z)} \approx \left(\frac{z_q}{z_p} \right) \frac{1}{L} \sum_{l=1}^L \tilde{r}_l f(z_l)$$

$$\tilde{r}_l = \frac{\tilde{p}(z_l)}{\tilde{q}(z_l)}$$

$$\frac{z_q}{z_p} \approx L \times \frac{1}{\sum_{l=1}^L \tilde{r}_l} \quad \text{for } z$$

$$\mathbb{E}_{z \sim p(z)} \approx \frac{1}{\sum_{l=1}^L \tilde{r}_l} \times \sum_{l=1}^L \tilde{r}_l f(z_l) = \sum_{l=1}^L w_l f(z_l)$$

$$\text{where } w_l = \frac{\tilde{r}_l}{\sum_{m=1}^M \tilde{r}_m} = \frac{\tilde{p}(z_l)/\tilde{q}(z_l)}{\sum_{m=1}^M \tilde{p}(z_m)/\tilde{q}(z_m)} \times z_q = \frac{\tilde{p}(z_l)/\tilde{q}(z_l)}{\sum_{m=1}^M \tilde{p}(z_m)/\tilde{q}(z_m)} \times z_q$$

Resampling)

1. L 個の sample

$z_1, \dots, z_L \in p(z)$ を抽出

2. 前1ステップの w_k を用いて $w_1, \dots, w_L$ を計算

3. 1つの sample を引いた後

2ステップの w_1 に基づいて L 個の sample

再抽出

これは $p(z)$ の近似に近づく

$L \rightarrow \infty$ 正確になる

MCMC

任意の分布 $p(x)$ と 不変分布に持つマルコフ連鎖を生成

初期値に依存しない

与えられた
乱数を生成

推移確率を与える手法

↳ メトロポリス法・ギブスサンプリング

ランダム walk は 状態空間 確率に 向かう

$\Rightarrow$ 現在の状態は (この) 状態にしか依存しない

$$p(x^{m+1} | x^1, \dots, x^m) = p(x^{m+1} | x^m)$$

↑
推移確率

エルゴード性 (この状態からどこまで

すべての状態にいく)

分布の定常 (不変) 性

この step にかいて $z' \rightarrow z$ の遷移確率
同じ

$$P^*(z) = \sum_{z'} T(z', z) P^*(z')$$

対称性 (可逆)

$$P^*(z) T(z, z') = P^*(z') T(z', z)$$

相関分布の z に 関与する 分布から成るたい

不変分布 $P^*(z)$ に 収束する z である

Metropolis - Hastings Algo



→ Metropolis Algo - 非変性

→ 変換する
しついで
(収束
する)

この特別 ver の Gibbs Sampling

