

$$L(\theta) := -\sum_n \log P_\theta(y_n | x_n)$$

[1]

本来的 Fisher

$$P_\theta(y|x) = P(y | f(x; \theta)) \quad \theta = \theta$$

$$\bar{F} \Pi_n P_\theta(x, y) = N \#_{x, y \sim P(x) P(y|x)} [\dots]$$

↳ FIM of model $P_\theta(x, y)$

$$[\dots] = \left[\mathbb{E}_\theta \log P_\theta(y|x) \mathbb{E}_\theta \log P_\theta(y|x) \right]$$

$P(x)$ is "input (= 入力) distribution" $P(y|x)$ is "conditional distribution"

FIM of conditional distribution $P(y|x)$

$$\Pi_n P_\theta(y|x_n) \text{ を使う}$$

[2] ML の true fisher

$$\bar{F} \Pi_n P_\theta(y|x_n) = \sum_n \#_{y \sim P_\theta(y|x_n)} [\dots]$$

$$[\dots] = \left[\mathbb{E}_\theta \log P_\theta(y|x_n) \mathbb{E}_\theta \log P_\theta(y|x_n) \right]$$

↳ (= Empirical) $\theta = \hat{\theta}$ を使う = $\theta = \hat{\theta}$

[3] ML での Empirical Fisher

$$\tilde{F} = \sum_n \left[\nabla_{\theta} \log p_{\theta}(y_n | x_n) \nabla_{\theta} \log p_{\theta}(y_n | x_n) \right]$$

$x \in y \in$ Empirical Data である。

$$= \sum_n \mathbb{E}_{y \sim P(y|x_n)} \left[\nabla_{\theta} \log p_{\theta}(y_n | x_n) \nabla_{\theta} \log p_{\theta}(y_n | x_n) \right]$$

$y \sim P(y|x_n)$ である。これは $\sum_{y \in \mathcal{Y}}$

$P(y|x_n)$ は y の分布
 $P_{\theta}(y|x_n)$ は θ による分布

Cross Entropy

$$H(q, p)$$

q の y に対する
 p の predict

$$= \sum_d \sum_c \text{data class } q_{dc} \log p_{dc}$$

ミニマムの導出

Chaoqi の KFAc - PyTorch

$$L \Rightarrow -\sum_n \log p_\theta(y_n | x_n)$$

の最小化をやる

なぜなら

ML の目標 - $p_\theta(y|x)$ を 真の分布 $q(y|x)$ に近づきたい

KL 距離は

$$D_{KL}(q|p_\theta) = \mathbb{E}_{y \sim q(y|x)} [\log q(y|x) - \log p_\theta(y|x)]$$

$$= \underbrace{\mathbb{E}_{y \sim q(y|x)} [\log q(y|x)]}_{\text{定数である}(\theta \text{ に依存しない)}} - \underbrace{\mathbb{E}_{y \sim q(y|x)} [\log p_\theta(y|x)]}_{\text{cross-entropy}}$$

定数である (θ に依存しない)

↑
cross-entropy
に相当

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} D_{KL}(q|p_\theta)$$

$$= \underset{\theta}{\operatorname{argmin}} \mathbb{E}_{y \sim p(y|x)} [-p_\theta \log(q|x)]$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \mathbb{E}_{y \sim q(y|x)} [-\log p_{\theta}(y|x)]$$

これに対し x を使ったものはいいのか？

↳ Ans. x を ~~確率分布~~ 空間に定めた生成モデル
考えよう

周辺分布

$q(x)$
を定めた

x を定めた ~~生成モデル~~ cross entropy
の値を考えよう

$$\mathbb{E}_{x \sim q(x)} \left[\mathbb{E}_{y \sim q(y|x)} [-\log p_{\theta}(y|x)] \right]$$

$$= \mathbb{E}_{x, y \sim q(x, y)} [-\log p_{\theta}(y|x)]$$

最終的に $q(x, y)$ の F_2 の $-\log p_{\theta}(y|x)$

の期待値の最小値
をとり

よって $g(x, y)$ は $\hat{g}(x, y)$ である

これは Empirical Distribution
である

MLとは

$$L(\theta) = \mathbb{E}_{x, y \sim \hat{g}(x, y)} [-\log p_{\theta}(y|x)]$$

これは $D_{KL}(g(y|x) | p_{\theta}(y|x))$ の

最小値

⇓ ($\log g$ は constant である)

Cross entropy の最小値

⇓

$$\mathbb{E}_{x, y \sim \hat{g}(x, y)} [\log p_{\theta}(y|x)]$$

各条件 x での $p_{\theta}(y|x)$ の最小値 =
それ自身

$$= -\sum_u \log p_{\theta}(y_u|x_u)$$

$$\hat{\theta} = \hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmin}} L(\theta)$$

CE is cross entropy の ~~式~~

6

実装に注意度す

L156

$$y_n \sim P_\theta(y_n | x_n)$$

L157

$$\text{loss} = \text{CE}(p_n, y_n)$$

これは gradient

$\nabla_{\theta} \log p_{\theta}(y_n | x_n)$ を計算するのには
用いず

L162

$$\text{sample}_y \sim P_{\theta}(y | x_n)$$

multinomial 実装度す

L164

$$\text{loss-sample} = \text{CE}(p_n, \text{sample}_y)$$

$$\mathbb{E}_{y_n \sim P_{\theta}(y | x_n)} [\nabla_{\theta} \log p_{\theta}(y | x_n)]$$

を計算するのには用いず