# MINE: Mutual information Neural Estimation

山田真徳

# 目的：相互情報量をニューラルネットから計算する

背景：相互情報量は重要な量(例えばVAEとか)なのに計算するのが大変。
離散変数か確率分布が既知じゃないと計算できないし、ノンパラメトリックカーネル密度推定やガウス性を仮定しても高次元になると計算が大変

戦略：KL diverenceをDonsker-Varahan 表現で書き直しその下限を最大化する

# MINE導出の流れ

KL-divergenceの下限を書き直す $D_{KL}(\mathbb{P}\,||\,\mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$

相互情報量の場合に書き直す $I_\Theta(X,Z) = \sup_{\theta\in\Theta} \mathbb{E}_{P(X,Z)}[T_\theta] - \log(\mathbb{E}_{P(X)P(Z)}[e^{T_\theta}])$

Tをニューラルネットで表現し最大化する

**Algorithm 1** . Mutual Information Estimation

$\theta \leftarrow$ initialize network parameters
**repeat**
  $(x^{(1)}, z^{(1)}), \ldots, (x^{(n)}, z^{(n)}) \sim \mathbb{P}_{XZ}$  ▷ Draw $n$ samples from the joint distribution
  $\bar{z}^{(1)}, \ldots, \bar{z}^{(n)} \sim \mathbb{P}_Z$  ▷ Draw $n$ samples from the $Z$ marginal distribution
  $\mathcal{V}(\theta) \leftarrow \frac{1}{n}\sum_{i=1}^n T_\theta(x^{(i)}, z^{(i)}) - \log(\frac{1}{n}\sum_{i=1}^n e^{T_\theta(x^{(i)}, \bar{z}^{(i)})})$
  ▷ Evaluate the lower-bound
  $\theta \leftarrow \theta + \nabla_\theta \mathcal{V}(\theta)$  ▷ Update the statistic network parameters
**until** convergence

## 定理１：KLは以下の表現を持つ(Donsker-Varahan representation)

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = \sup_{T:\Omega \to \mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^{T}])$$

Ω：measure space
T ：function

## 証明

Boltzmann distribution $\mathbb{G}$ を考える

$$d\mathbb{G} = \frac{1}{Z} e^T d\mathbb{Q} \qquad Z = \mathbb{E}_{\mathbb{Q}}[e^T]$$

以下が成り立つ

$$d\mathbb{G} = \frac{1}{Z} e^T d\mathbb{Q}$$

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[T] - \log Z &= \mathbb{E}_{\mathbb{P}}[\log e^T] - \log(e^T \frac{d\mathbb{Q}}{d\mathbb{G}}) \\
&= \mathbb{E}_{\mathbb{P}}\left[\log e^T - \log e^T - \log \frac{d\mathbb{Q}}{d\mathbb{G}}\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{G}}{d\mathbb{Q}}\right] \qquad \cdots ①
\end{aligned}$$

以下を示すことで証明する

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = \Delta + \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T]) \geq \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

Δは最初の等式が成り立つように定義する　　Δ≧0を示せば証明終了

$$\Delta \equiv D_{KL}(\mathbb{P}||\mathbb{Q}) - \left(\mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])\right) \geq 0$$

$$Z = \mathbb{E}_{\mathbb{Q}}[e^T]$$

$$\mathbb{E}_{\mathbb{P}}[T] - \log Z = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{G}}{d\mathbb{Q}}\right] \quad \cdot\cdot\cdot ①$$

$$\Delta = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}} - \log \frac{d\mathbb{G}}{d\mathbb{Q}}\right] = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{G}}\right] = D_{KL}(\mathbb{P}||\mathbb{G})$$

KLなので正

証明終了

等式の条件

Δ=0 ↔ P=Q

$$T^* = \log \frac{d\mathbb{P}}{d\mathbb{G}} + C$$

Cは定数

$$T = \log \frac{d\mathbb{P}}{d\mathbb{Q}} + \log \mathbb{E}_{\mathbb{Q}}[e^T]$$

$$T^* = \log \frac{d\mathbb{P}}{d\mathbb{Q}} + C$$

# 相互情報量のときのDonsker-Varahan representation

DNNで表現された関数族 $T_\theta : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ で相互情報量を近似する

$$I_\Theta(X, Z) = \sup_{\theta \in \Theta} \mathbb{E}_{P(X,Z)}[T_\theta] - \log(\mathbb{E}_{P(X)P(Z)}[e^{T_\theta}])$$

実際はp(x,z)とp(x)とp(z)からのモンテカルロサンプリングで計算される

$$\widehat{I(X; Z)}_n = \sup_{\theta \in \Theta} \mathbb{E}_{P^{(n)}(X,Z)}[T_\theta] - \log(\mathbb{E}_{P^{(n)}(X)\hat{P}^{(n)}(Z)}[e^{T_\theta}])$$

**Algorithm 1** . Mutual Information Estimation

$\theta \leftarrow$ initialize network parameters
**repeat**
  $(x^{(1)}, z^{(1)}), \ldots, (x^{(n)}, z^{(n)}) \sim \mathbb{P}_{XZ}$ ▷ Draw $n$ samples from the joint distribution
  $\bar{z}^{(1)}, \ldots, \bar{z}^{(n)} \sim \mathbb{P}_Z$ ▷ Draw $n$ samples from the $Z$ marginal distribution
  $\mathcal{V}(\theta) \leftarrow \frac{1}{n}\sum_{i=1}^{n} T_\theta(x^{(i)}, z^{(i)}) - \log(\frac{1}{n}\sum_{i=1}^{n} e^{T_\theta(x^{(i)}, \bar{z}^{(i)})})$
  ▷ Evaluate the lower-bound
  $\theta \leftarrow \theta + \nabla_\theta \mathcal{V}(\theta)$ ▷ Update the statistic network parameters
**until** convergence