統計的機械学習
第九回　レポート ID: 02

37-196360　森田涼介

2019 年 6 月 12 日

以下の，周辺尤度と KL 情報量の関係を導出する。

$$\log p(x_{1:n}|\eta) - L[q(z_{1:n})q(\theta); x_{1:n}] = \mathrm{KL}[q(z_{1:n})q(\theta)|p(z_{1:n}, \theta|x_{1:n}, \eta)] \tag{1}$$

データ $x_{1:n}$ の生成確率の周辺尤度からその変分下限を得る。

$$\begin{aligned}
\log p(x_{1:n}|\eta) &= \log \int p(x_{1:n}, z_{1:n}, \theta|\eta)\, \mathrm{d}z_{1:n}\, \mathrm{d}\theta && (2)\\
&= \log \int q(z_{1:n})q(\theta) \frac{p(x_{1:n}, z_{1:n}, \theta|\eta)}{q(z_{1:n})q(\theta)}\, \mathrm{d}z_{1:n}\, \mathrm{d}\theta && (3)\\
&\geq \int q(z_{1:n})q(\theta) \log \frac{p(x_{1:n}, z_{1:n}, \theta|\eta)}{q(z_{1:n})q(\theta)}\, \mathrm{d}z_{1:n}\, \mathrm{d}\theta && (4)\\
&\equiv L[q(z_{1:n})q(\theta); x_{1:n}] && (5)
\end{aligned}$$

また，KL 情報量（Kullback-Leibler Divergence）は一般に，

$$\mathrm{KL}[q(\theta)|p(\theta)] = \int q(\theta) \log \frac{q(\theta)}{p(\theta)} \mathrm{d}\theta \tag{6}$$

と表されるから，

$$\mathrm{KL}[q(z_{1:n})q(\theta)|p(z_{1:n}, \theta|x_{1:n}, \eta)] = \int q(z_{1:n})q(\theta) \log \frac{q(z_{1:n})q(\theta)}{p(z_{1:n}, \theta|x_{1:n}, \eta)}\, \mathrm{d}z_{1:n}\, \mathrm{d}\theta \tag{7}$$

である。ここで，

$$\begin{aligned}
\frac{p(x_{1:n}, z_{1:n}, \theta|\eta)}{p(z_{1:n}, \theta|x_{1:n}, \eta)} &= \frac{p(x_{1:n},\ z_{1:n},\ \theta,\ \eta)}{p(\eta)} \cdot \frac{p(x_{1:n},\ \eta)}{p(x_{1:n},\ z_{1:n},\ \theta,\ \eta)} && (8)\\
&= \frac{p(x_{1:n},\ \eta)}{p(\eta)} && (9)\\
&= p(x_{1:n}|\eta) && (10)
\end{aligned}$$

であることに注意すると，変分下限と KL 情報量の和は，

$$
\begin{aligned}
&L[q(z_{1:n})q(\theta);x_{1:n}]+\mathrm{KL}[q(z_{1:n})q(\theta)|p(z_{1:n},\theta|x_{1:n},\eta)] && (11)\\
&\quad=\int q(z_{1:n})q(\theta)\log\frac{p(x_{1:n},z_{1:n},\theta|\eta)}{q(z_{1:n})q(\theta)}\,\mathrm{d}z_{1:n}\,\mathrm{d}\theta+\int q(z_{1:n})q(\theta)\log\frac{q(z_{1:n})q(\theta)}{p(z_{1:n},\theta|x_{1:n},\eta)}\,\mathrm{d}z_{1:n}\,\mathrm{d}\theta && (12)\\
&\quad=\int q(z_{1:n})q(\theta)\log\frac{p(x_{1:n},z_{1:n},\theta|\eta)}{p(z_{1:n},\theta|x_{1:n},\eta)}\,\mathrm{d}z_{1:n}\,\mathrm{d}\theta && (13)\\
&\quad=\int q(z_{1:n})q(\theta)\log p(x_{1:n}|\eta)\,\mathrm{d}z_{1:n}\,\mathrm{d}\theta && (14)
\end{aligned}
$$

となる。いま，$q(z_{1:n})$, $q(\theta)$ は確率密度より

$$\int q(z_{1:n})\,\mathrm{d}z_{1:n}=\int q(\theta)\,\mathrm{d}\theta=1 \tag{15}$$

であること，及び $\log p(x_{1:n}|\eta)$ は $z_{1:n}$, θ に依存しないことから，結局，

$$L[q(z_{1:n})q(\theta);x_{1:n}]+\mathrm{KL}[q(z_{1:n})q(\theta)|p(z_{1:n},\theta|x_{1:n},\eta)]=\log p(x_{1:n}|\eta) \tag{16}$$

よって，

$$\log p(x_{1:n}|\eta)-L[q(z_{1:n})q(\theta);x_{1:n}]=\mathrm{KL}[q(z_{1:n})q(\theta)|p(z_{1:n},\theta|x_{1:n},\eta)] \tag{17}$$