Variational Auto Encoder

jojonki

January 2018

1　導出するよ

Kingma氏のVAEの論文で(https://arxiv.org/pdf/1312.6114.pdf)，Appendix Bに多変量正規分布間におけるKLダイバージェンスの計算の式が載っていたが，途中式が省かれていたので導出を試みた.

変分下限の正則化項の導出を目標にこの資料を作ったが，せっかくなので大本の変分下限の下りも少しだけ触れる． $p(z|x)$の近似分布として，$q(z|x)$を考え，$D_{KL}(q(z|x)||p(z|x))$を考えるわけだが，変分下限を利用すると，

$$
\begin{aligned}
D_{KL}(q(z|x)||p(z|x)) &= \int q(z|x) \log \frac{q(z|x)}{p(z|x)} dz \\
&= \int q(z|x) \Big\{ \log q(z|x) - \log p(z|x) \Big\} dz \\
&\quad p(z|x)\text{をベイズの定理で展開} \\
&= \int q(z|x) \Big\{ \log q(z|x) - \log \frac{p(x|z)p(z)}{p(x)} \Big\} dz \\
&= \int q(z|x) \Big\{ \log q(z|x) - \log p(x|z) - \log p(z) + \log p(x) \Big\} dz \\
&= \int q(z|x) \log p(x) dz + \int q(z|x) \Big\{ \log q(z|x) - \log p(x|z) - \log p(z) \Big\} dz \\
&\quad \text{第 1 項は周辺化で積分が消える} \\
&= \log p(x) + \int q(z|x) \Big\{ \log q(z|x) - \log p(x|z) - \log p(z) \Big\} dz \\
&= \log p(x) + \int q(z|x) \Big\{ \log \frac{q(z|x)}{p(z)} - \log p(x|z) \Big\} dz \\
&= \log p(x) + \int q(z|x) \log \frac{q(z|x)}{p(z)} dz - \int q(z|x) \log p(x|z) dz \\
&\quad \text{第 2 項はKLダイバージェンスの定義そのもの} \\
&= \log p(x) + D_{KL}(q(z|x)||p(z)) - \int q(z|x) \log p(x|z) dz \\
&= \log p(x) + D_{KL}(q(z|x)||p(z)) - \mathbb{E}_{q(z|x)}[\log p(x|z)]
\end{aligned} \tag{1}
$$

これを下記のように並べ替えてみる.

$$
\log p(x) - D_{KL}(q(z|x)||p(z)) = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(z)) \tag{2}
$$

KLダイバージェンスは0以上であるわけだから，左辺のKLダイバージェンスを取り除くと下記の不等号が

成り立つ．

$$\log p(x) \geq \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(z)) \tag{3}$$

このときの右辺がEvidence Lower BOund (ELBO)と呼ばれ，これを下限を最大化することで$p(x)$を高めることができる．第1項が復元誤差(Reconstruction error)，第2項が正則化項となる．

$$\mathcal{L} = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(x|z)) \tag{4}$$

1.1 復元誤差の導出

復元誤差の部分は，難しくない．下記のように近似することでL個サンプリングしてその平均をとる．（1個とかでもMNISTはうまくいっているっぽい)．

$$\mathbb{E}_{q(z|x)}[\log p(x|z)] \simeq \frac{1}{L}\sum_{l=1}^{L}\log p(x|z) \tag{5}$$

MNIST等の画像を考え，各ピクセルの値は0-1とするとベルヌーイ分布を想定できる．つまりBinary Cross Entropyが使える．Dはピクセル数．

$$\log p(x|z) = \sum_{i=1}^{D}\left\{x_i \log(x_i) + (1-x_i)\log(1-x_i)\right\} \tag{6}$$

1.2 正則化項の導出

正則化項は正規分布間のKLダイバージェンスを求めるので式が長いのだけど，やっていることは単純であるため，一度手計算をしてみると良い．

$$\begin{aligned} D_{KL}(q(z|x)||p(z)) &= \int q(z|x)\log\frac{q(z|x)}{p(z)}dz \\ &= \int q(z|x)\{\log q(z|x) - \log p(z)\}dz \\ &= \int q(z|x)\log q(z|x)dz - \int q(z|x)\log p(z)dz \end{aligned} \tag{7}$$

第 1 項と第 2 項をそれぞれ求めたいと思う． $p(z)$はVAEの設定から $\sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.であり，$q(z|x)$は$\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$となる．

第1項

$$
\begin{aligned}
\int q(z|x)\log q(z|x)dz &= \int \mathcal{N}(z;\mu,\sigma^2)\log\mathcal{N}(z;\mu,\sigma^2)dz \\
&= \sum_j^J E_{q(z_j)}\Big[\log\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\big(-\frac{(z_j-\mu_j)^2}{2\sigma_j^2}\big)\Big] \\
&= \sum_j^J E_{q(z_j)}\Big[-\frac{1}{2}\log(2\pi\sigma_j^2)-\frac{(z_j-\mu_j)^2}{2\sigma_j^2}\Big] \\
&= \sum_j^J \Big\{-\frac{1}{2}\log(2\pi\sigma_j^2)-E_{q(z_j)}\Big[\frac{(z_j-\mu_j)^2}{2\sigma_j^2}\Big]\Big\} \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J\log\sigma_j^2-\frac{1}{2}\sum_{j=1}^J E_{q(z_j)}\Big[\frac{(z_j-\mu_j)^2}{\sigma_j^2}\Big] \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J\log\sigma_j^2-\frac{1}{2}\sum_{j=1}^J \frac{1}{\sigma_j^2}E_{q(z_j)}\Big[(z_j-\mu_j)^2\Big] \\
&\quad \text{第3項の右側は分散の定義そのもの.} \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J\log\sigma_j^2-\frac{1}{2}\sum_{j=1}^J \frac{1}{\sigma_j^2}\sigma_j^2 \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J(\log\sigma_j^2+1)
\end{aligned} \tag{8}
$$

第2項

$$
\begin{aligned}
\int q(z|x)\log p(z)dz &= \int \mathcal{N}(z;\mu,\sigma^2)\log\mathcal{N}(z;\mathbf{0},\boldsymbol{I})dz \\
&= \sum_j^J E_{q(z_j)}\Big[\log\frac{1}{\sqrt{2\pi}}\exp\big(-\frac{z_j^2}{2}\big)\Big] \\
&= \sum_j^J E_{q(z_j)}\Big[-\frac{1}{2}\log(2\pi)-\frac{z_j^2}{2}\Big] \\
&= \sum_j^J \Big\{-\frac{1}{2}\log(2\pi)-E_{q(z_j)}\Big[\frac{z_j^2}{2}\Big]\Big\} \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J E_{q(z_j)}\Big[z_j^2\Big] \\
&\quad \text{ここで第２項に分散の公式を使う.} \quad \sigma^2 = E[X^2]-\mu^2 \\
&= -\frac{J}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^J(\sigma_j^2+\mu_j^2)
\end{aligned} \tag{9}
$$

式(7)の第 1 項と第 2 項が求まったので，式(7)をまとめると下記のようになり，論文の結果と一致することになる．(論文ではKLダイバージェンスに負の符号がついているので注意).

$$
\begin{aligned}
D_{KL}(q(z|x)||p(z)) &= \int q(z|x)\log\frac{q(z|x)}{p(z)}dz \\
&= \int q(z|x)\{\log q(z|x) - \log p(z)\}dz \\
&= \int q(z|x)\log q(z|x)dz - \int q(z|x)\log p(z)dz \\
&= \left(-\frac{J}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{J}(\log\sigma_j^2 + 1)\right) - \left(-\frac{J}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{J}(\sigma_j^2 + \mu_j^2)\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\left(1 + \log\sigma^2 - \sigma_j^2 - \mu_j^2\right)
\end{aligned}
\tag{10}
$$

2 参考

- 元論文 `https://arxiv.org/pdf/1703.10960.pdf`
- 猫でも分かるVariational AutoEncoder．概念の理解に．`https://www.slideshare.net/ssusere55c63/variational-autoencoder-64515581`
- nzwさんによる導出．数式の理解に．`https://nzw0301.github.io/notes/vae.pdf`