# 変分ベイズ法からVAEへ

持橋大地
統計数理研究所 数理・推論研究系
daichi@ism.ac.jp
首都大小町研究室講義, 2018-2-6(火)

# 変分ベイズで何が不足？

- 共役分布族 (ディリクレ-多項分布など)でないとEMアルゴリズムが導けない

期待値 $\left\langle \log p(D, z|\theta) \right\rangle_{q(z)}, \left\langle \log p(D, z|\theta) \right\rangle_{q(\theta)}$

が解析的に解けない

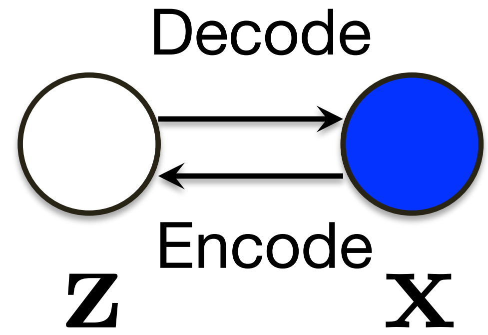
- 強い因子分解の仮定

$$q(z, \theta) = q(z)q(\theta)$$

- 複雑なデータの正確なモデル化が難しい

# 逆に、ニューラルネットでは？

- 通常のオートエンコーダでは、新しいデータを生成できない
  - 潜在変数 z に分布がない
- きちんとした確率的生成モデルになっていない！

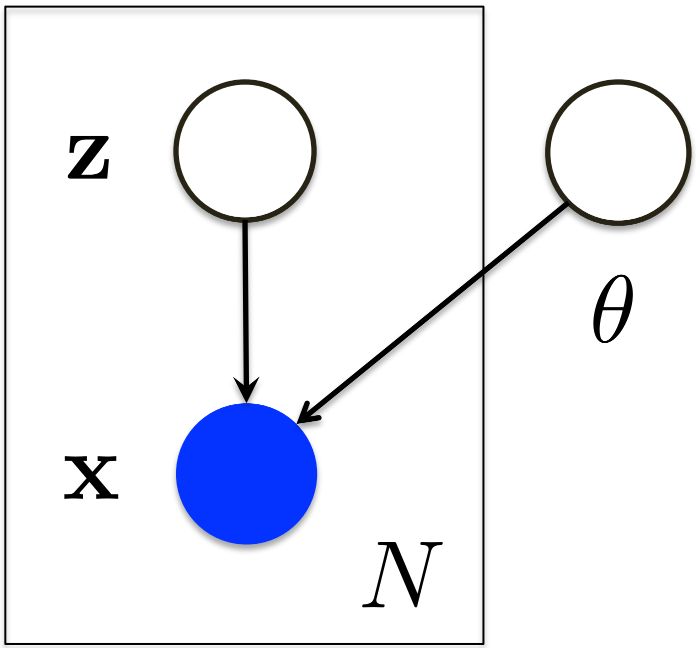


# Variational Autoencoder (VAE)

- Kingma & Welling (ICLR 2014)
  - 元のタイトル: “*Stochastic Gradient VB and the Variational Auto-Encoder*”
- 変分下限をニューラルネットで近似
  - 因子分解不要
  - Stochastic Gradientで学習可能
  - z は典型的には多変量標準ガウス分布

$$\mathbf{z} \sim \mathbb{N}(\mathbf{0}, \mathbf{I}_K)$$

# VAEのモデル: 前提




- θの分布は(とりあえず)考えない
(定数として最適化)
- $p(\mathbf{x}|\theta) = \int p(\mathbf{x}|\mathbf{z},\theta)p(\mathbf{z})d\mathbf{z}$
を最大化
- $\mathbf{z} \sim \mathbb{N}(\mathbf{0}, \mathbf{I}_K)$ が典型的

# VAEの導出 (1)

$$\log p(\mathbf{x}|\theta) = \log \int p(\mathbf{x},\mathbf{z}|\theta) d\mathbf{z} = \log \int q(\mathbf{z}|\phi) \frac{p(\mathbf{x},\mathbf{z}|\theta)}{q(\mathbf{z}|\phi)} d\mathbf{z}$$

$$\geq \int q(\mathbf{z}|\phi) \log \frac{p(\mathbf{x},\mathbf{z}|\theta)}{q(\mathbf{z}|\phi)} d\mathbf{z} \qquad \text{(Jensen)}$$

$$= \int q(\mathbf{z}|\phi) \log \frac{p(\mathbf{x}|\mathbf{z},\theta)p(\mathbf{z}|\theta)}{q(\mathbf{z}|\phi)} d\mathbf{z}$$

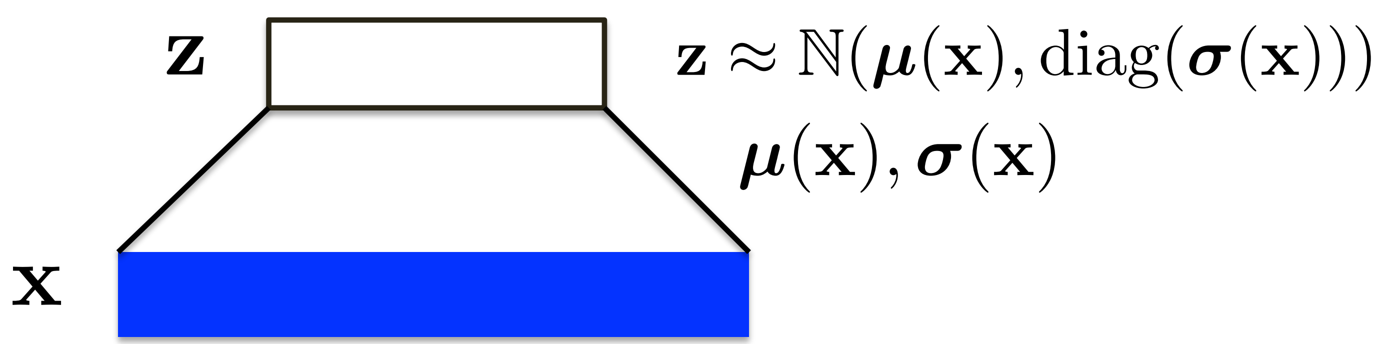
$$= \underbrace{-D(q(\mathbf{z}|\phi)||p(\mathbf{z}|\theta))}_{(1)} + \underbrace{\int q(\mathbf{z}|\phi) \log p(\mathbf{x}|\mathbf{z},\theta) d\mathbf{z}}_{(2)}$$

- Jensenの不等式で下限を取っているだけ
- (2)項を最大化したいが、(1)項がペナルティ(正則化)

# VAEの導出 (2)

$$\log p(\mathbf{x}|\theta) \geq \underbrace{-D(q(\mathbf{z}|\phi)||p(\mathbf{z}|\theta))}_{(1)} + \underbrace{\int q(\mathbf{z}|\phi) \log p(\mathbf{x}|\mathbf{z},\theta) d\mathbf{z}}_{(2)}$$

- $q(\mathbf{z}|\phi) = \mathbb{N}(\boldsymbol{\mu}(\phi), \mathrm{diag}(\boldsymbol{\sigma}(\phi)))$ とすれば、
  (1) は解析的
  (2) はモンテカルロ近似できる
  } SGDで最適化できる!



# VAEの導出 (3)

- 正規分布の間のKLダイバージェンス D(q(z|φ)||p(z)) は解析的に求められるので、結局

$$\log p(\mathbf{x}|\theta) \geq \frac{1}{2}\sum_{k=1}^{K}(1+\log\sigma_k^2-\mu_k^2-\sigma_k^2)+\frac{1}{L}\sum_{\ell=1}^{L}\log p(\mathbf{x}|\mathbf{z}^{(\ell)},\theta)$$

$$\mathbf{z}^{(\ell)} \sim q(\mathbf{z}|\phi)$$

  – この下限のGradientを計算してSGDに入れればよい.
  – Reparametrization trick (正規乱数の変換)

# VAE: ポイント

- Jensenで下限をとった後のq(z|φ)は任意の関数でよい
  - 因子化仮定は必須ではない
  - q(z|φ)をニューラルネットでモデル化
- ガウス分布間のKLダイバージェンスは解析的に計算できる
  - 本質的には必ずガウス分布である必要はない

VAE=変分近似＋ニューラルネット
ニューラルネットの確率的生成モデル化.

# VAE: 注意点

ただし...

- Jensenの不等式により、D(q(z|ϕ)||p(z)) を最小化
  × D(p(z)||q(z|ϕ)) ではない

$$D(q(\mathbf{z}|\phi)||p(\mathbf{z})) = \int \boxed{q(\mathbf{z}|\phi)} \log \frac{q(\mathbf{z}|\phi)}{p(\mathbf{z})} d\mathbf{z}$$

  - モデル化している範囲で、q(z|ϕ)とp(z)の差が小さければよい
  - p(z)が大, q(z|ϕ)が小の領域があっても気にしない
  - Peakyな事後分布 (Mode-finding: PRML参照)

# まとめ

- VAE ... 変分ベイズ法の正統進化
  - q(z|φ)として解析関数ではなく、p(z)と同じ分布族(たとえば正規分布)を与えるニューラルネットを考える
  - 勾配が解析的には解けないので、モンテカルロ近似
- 通常のNNと異なり、データを簡単に生成できる
- 変分下限を考える際に、真の事後分布より尖っている可能性が高いので注意が必要