Tutorial on Variational Autoencoders

清水　豪士

富山県立大学　情報基盤工学講座
t715038@st.pu-toyama.ac.jp

June 11, 2021

はじめに

背景

- Variational Autoencvorders：VAE は複雑な分布を教師なしで学習するための最も一般的なアプローチの１つである.

目的

- VAE のいくつかの動作を紹介する.

VAE

- バックプロパゲーションに基づく関数近似器を使用して生成モデルを構築できる有望なフレームワーク.
- バックプロバゲーションにより高速に学習できる.

潜在変数モデル

潜在変数モデル

- 生成モデルを学習する際，次元間の依存関係が複雑であればあるほど，モデルの学習は難しくなる．
- 高次元空間 Z に潜在変数 z のベクトルがあり，Z 上で定義された確率密度関数 (PDF)$P(z)$ にしたがって簡単にサンプリングできるとする．
- $P(z)$ から z をサンプリングように θ を最適化すると，$f(z;\theta)$ がデータセットの X のようになる．
- この概念を正確にするために，次に従って生成プロセス全体で学習セットの各 X の確率を最大化する．

$$P(X) = \int P(X|z;\theta)P(z)dz. \tag{1}$$

$f(z;\theta)$ は，分布 $P(X|z;\theta)$ に置き換える．
これにより，全体確率の法則を用いて，X の z への依存度を明示することができる．
（以降，$f(z;\theta)$ から θ を省略する）

VAE では，出力分布の選択はガウス，つまり，
$P(X,z;\theta) = N(X|f(z;\theta),\sigma^2 * I)$ となる．ガウス分布をもつことで，勾配

降下法を用いて，ある z に対して $f(z;\theta)$ を X に近づけることで $P(X)$ を増加させることができる．
つまり，生成モデルの下で学習データを徐々に高めていくことができる．

変分オートエンコーダ

- VAE は古典的なオートエンコーダとは，ほとんど関係ない．
- VAE は $P(X)$ から直接サンプリングできる．

式 1 を計算するための変分オートエンコーダの重要なアイデアは，X を生成した可能性の高い z の値をサンプリングし，そこから $P(X)$ を計算すること．
つまり，X の値を受け取って，X を生成する可能性の高い z の値の分布を与えることができる関数 $Q(z|X)$ が必要になる．

変分オートエンコーダー

$P(z|X)$ と $Q(z)$ の間の KL divergence をある任意の Q について定義する.

$$\mathcal{D}\left[Q(z)\|P(z|X)\right] = E_{z\sim Q}\left[\log Q(z) - \log P(z|X)\right]. \tag{2}$$

$P(z|X)$ にベイズ則を適用することで $P(X)$ と $P(X|z)$ の両方を式にいれることができる.

$$\mathcal{D}\left[Q(z)\|P(z|X)\right] = E_{z\sim Q}\left[\log Q(z) - \log P(X|z) - \log P(z)\right] + \log P(X). \tag{3}$$

ここで, $logP(X)$ は, z に依存しない.
並べ替えて, $E_{z\sim Q}$ の一部を KL-divergence の項に縮めると

$$\log P(X) - \mathcal{D}\left[Q(z|X)\|P(z|X)\right] = E_{z\sim Q}\left[\log P(X|z)\right] - \mathcal{D}\left[Q(z|X)\|P(z)\right]. \tag{5}$$

この式が変分オートエンコーダのコアである.
左辺には最大化したい $logP(X)$ がある.
右辺には Q が正しく選択された際に, 確率的勾配降下法によって最適化することができる.

式 5 の $D[Q(z|X)||P(z)]$ は，2 つの多変量ガウス分布の間の KLdivergence であり，閉形式で次のように計算できる．

$$\mathcal{D}[\mathcal{N}(\mu(X),\Sigma(X))\|\mathcal{N}(0,I)] = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma(X)\right) + \left(\mu(X)\right)^{\top}\left(\mu(X)\right) - k - \log\det\left(\Sigma(X)\right)\right). \quad (7)$$

このとき，k は分布の次元数．μ，Σ はデータから学習可能なパラメータを持つ任意の決定論的関数．

式 5 の右辺の $E_{z\sim Q}[logP(X|z)]$ を推定するためには，コストがかかる．

そこで，確率的勾配降下法の標準的な方法として，z のサンプルを 1 つ取り，その z に対する $P(X|z)$ を $E_{z\sim Q}[logP(X|z)]$ の近似値として扱う．

最適化する方程式は

$$E_{X\sim D}\left[\log P(X) - \mathcal{D}\left[Q(z|X)\|P(z|X)\right]\right] = E_{X\sim D}\left[E_{z\sim Q}\left[\log P(X|z)\right] - \mathcal{D}\left[Q(z|X)\|P(z)\right]\right]. \tag{8}$$

分布 $Q(z|X)$ から z の単一の値をサンプリングして，次の式の勾配を計算することができる．

$$\log P(X|z) - \mathcal{D}\left[Q(z|X)\|P(z)\right]. \tag{9}$$

この関数の勾配を，X と z の任意の数のサンプルで平均かすると，式 8 の勾配に収束する．
$E_{z\backsim Q}[logP(X|z)]$ は P のパラメータだけでなく，Q のパラメータにも依存する．
しかし，式 9 にはこの依存性がなくなっている．
VAE を機能させるには，P が確実に解読できる X のコードを生成するように Q を動かすことが重要．

$Q(z|X)$ の平均と共分散である $\mu(X)$ と $\Sigma(X)$ があれば，$\epsilon \backsim N(0,1)$ をサンプリングし，$z = \mu(X) + \Sigma^{\frac{1}{2}}(X) * \varepsilon$ を計算することで，$N(\mu(X), \Sigma(X))$ からサンプリングできる．

実際に勾配をとる方程式は次のようになる．

$$E_{X\sim D}\left[E_{\epsilon\sim\mathcal{N}(0,I)}[\log P(X|z = \mu(X) + \Sigma^{1/2}(X) * \epsilon)] - \mathcal{D}\left[Q(z|X)\|P(z)\right]\right]. \tag{10}$$

固定された X と ϵ があれば，この関数は P と Q のパラメータで決定論的かつ連続的であり，バックプロパゲーションは確率的勾配降下法に対応した勾配を計算することができる．

条件付き変動オートエンコーダ：CVAE

- 今まで紹介した計算を修正し，生成プロセスを単純に入力に条件付けするもの．
- 入力から出力へのマッピングが 1 対多数である問題に取り組むことができる．

入力 X と出力 Y を与えられた時，グランドトゥルースの確率を最大化するモデル $P(\Upsilon|X)$ をつくる
このときに潜在変数 $z \sim N(0,1)$ を導入してモデルを定義する．

$$P(Y|X) = \mathcal{N}(f(z,X), \sigma^2 * I). \tag{12}$$

このとき，f はデータから学習できる決定論的な関数

条件付き変動オートエンコーダ

このとき，式2，式3，式5は以下のように書き換える．

$$\mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = E_{z\sim Q(\cdot|Y,X)}\left[\log Q(z|Y,X) - \log P(z|Y,X)\right] \tag{13}$$

$$\begin{aligned}&\mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = \\ &\quad E_{z\sim Q(\cdot|Y,X)}\left[\log Q(z|Y,X) - \log P(Y|z,X) - \log P(z|X)\right] + \log P(Y|X)\end{aligned} \tag{14}$$

$$\begin{aligned}&\log P(Y|X) - \mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = \\ &\qquad E_{z\sim Q(\cdot|Y,X)}\left[\log P(Y|z,X)\right] - \mathcal{D}\left[Q(z|Y,X)\|P(z|X)\right]\end{aligned} \tag{15}$$

MNIST：変分オートエンコーダ

VAE フレームワークの分布学習能力を示すために，MNIST と呼ばれる書き数字画像の大規模なデータベースで変分オートエンコーダを学習させる．学習は 1 回で完了した．

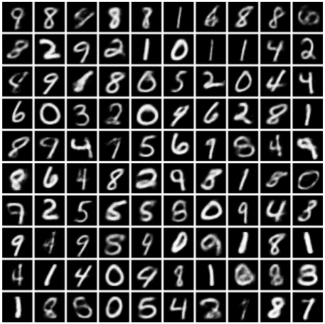

図 1: MNIST で学習した VAE モデル

かなりの数の数字が異なる数字の中間に位置している．
これは，連続的な分布を滑らかな関数でマッピングしているために起こる現象．

MNIST：条件付き変分オートエンコーダ

入力を数字の中央から取得した１列のピクセル.

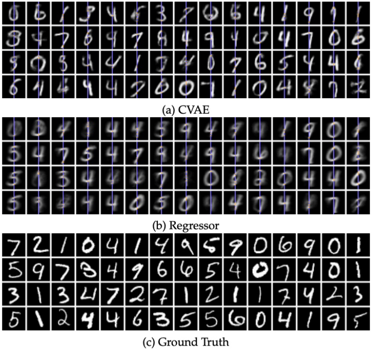

図 2: MNIST で学習した CVAE のサンプル

CVAE は，出力が曖昧な場合，リグレッサーよりも再現度を上回る.

リグレッサーでは出力をぼかすことで曖昧さを処理している.
CVAE は，ぼかしをかけずに出力するため，より信憑性の高い画像が得られる.