Diffusion models
Chapter1

矢口 真那斗

2024 年 6 月 7 日

目次

第 1 章

生成モデル

1.1　生成モデルとは何か

生成モデルとは、変数 θ でパラメトライズされた確率分布 $q_\theta(\mathbf{x})$ であり、この分布に従ってサンプリングすることとで、対象ドメインのデータを生成できるモデルである.
ここでは、生成モデルを訓練データから学習することを考える. 訓練データは iid である N 個のデータからなるデータセット $D = \{\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}\}$ であり、未知の確率分布 $p(\mathbf{x})$ に従うとする. ただし、$\mathbf{x}^{(i)} \in \mathbb{R}^d$ は d 次元ベクトルである i 番目のデータ、$x_i \in \mathbb{R}$ は $\mathbf{x}$ の i 次元目の成分である.
生成モデルの学習の目標は、目標確率分布 $p(\mathbf{x})$ にできるだけ近い確率分布 $q_\theta(\mathbf{x})$ を見つけることである.

1.2　エネルギーベースモデル・分配関数

データ $\mathbf{x} \in X$ に対して、エネルギー関数 $f(\mathbf{x};\theta) : \mathbb{R}^d \to \mathbb{R}$ を導入する. また生成モデルの確率分布 $q_\theta(\mathbf{x})$ を、エネルギー関数を用いて次のように定義する.

$$q_\theta(\mathbf{x}) = \frac{\exp(-f_\theta(\mathbf{x}))}{Z(\theta)} \tag{1.1}$$

$$Z(\theta) = \int_{\mathbf{x}' \in X} \exp(-f_\theta(\mathbf{x}'))d\mathbf{x}' \tag{1.2}$$

ここで、$Z(\theta)$ を分配関数とよぶ. このとき、エネルギー $f_\theta(\mathbf{x})$ が小さいほど、データ $\mathbf{x}$ が生成される確率が大きくなることを意味する. これは物理学におけるエネルギーの概念と対応しており、エネルギーベースモデルとよばれる.
エネルギーベースモデルの特徴として、構成性を挙げる. 2 つのエネルギー関数、$f_1(\mathbf{x}), f_2(\mathbf{x})$ と、対応する確率分布 $q_1(\mathbf{x}), q_2(\mathbf{x})$ を考える. このとき、エネルギー関

数の和 $f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ に対応する確率分布 $q(\mathbf{x})$ は、次のようになる.

$$\begin{aligned} q(\mathbf{x}) &\propto \exp(-f(\mathbf{x})) && (1.3)\\ &= \exp(-f_1(\mathbf{x}) - f_2(\mathbf{x})) && (1.4)\\ &= \exp(-f_1(\mathbf{x}))\exp(-f_2(\mathbf{x})) && (1.5)\\ &= q_1(\mathbf{x})q_2(\mathbf{x}) && (1.6)\end{aligned}$$

このように、エネルギー関数の和に対応する確率分布は、それぞれの確率分布の積になる.この性質は、生成モデルの構築において、複数のエネルギー関数を組み合わせる際に有用である.

1.3　学習手法

生成モデルの学習手法として、尤度ベースモデルとよばれる手法と、暗黙的生成モデルとよばれる手法の 2 つを紹介する.

1.3.1　尤度ベースモデル

あるデータ $\mathbf{x}$ の生成確率 $q_\theta(\mathbf{x})$ を尤度とよぶ. 訓練データセット $D = \{\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}\}$ の尤度は、データが i.i.d であるから次のようになる.

$$q_\theta(D) = q_\theta(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = \prod_{i=1}^{N} q_\theta(\mathbf{x}^{(i)}) \tag{1.7}$$

対数尤度 $L(\theta)$ は、次のようになる.

$$L(\theta) = \mathrm{log} q_\theta(D) = \sum_{i=1}^{N} \mathrm{log} q_\theta(\mathbf{x}^{(i)}) \tag{1.8}$$

対数尤度を最大にするパラメータ $\theta^*_{ML} := \arg\max_\theta L(\theta)$ を求めるのが、最尤推定である.

エネルギーベースモデルの場合の対数尤度を考える. $q_\theta(\mathbf{x}) = \frac{\exp(-f_\theta(\mathbf{x}))}{Z(\theta)}$ より、対数尤度は次のようになる.

$$\begin{aligned} L(\theta) &= \sum_{i=1}^{N} \mathrm{log} q_\theta(\mathbf{x}^{(i)}) && (1.9)\\ &= \sum_{i=1}^{N} \mathrm{log} \frac{\exp(-f_\theta(\mathbf{x}^{(i)}))}{Z(\theta)} && (1.10)\\ &= -\sum_{i=1}^{N} f_\theta(\mathbf{x}^{i}) - N\mathrm{log} Z(\theta) && (1.11)\\ &= -\sum_{i=1}^{N} f_\theta(\mathbf{x}^{i}) - N\mathrm{log} \int_{\mathbf{x}' \in X} \exp(-f_\theta(\mathbf{x}'))d\mathbf{x}' && (1.12)\end{aligned}$$

式 (1.12) より、対数尤度を最大にすることは、訓練データの位置のエネルギーを小さくし (第一項)、一方ですべての位置のエネルギーを大きくすることに対応する (第二項).
”$f_\theta(\mathbf{x}) = \exp(g_\theta(\mathbf{x}))$ と表せる場合を考えてみたい, 鋭くなる、記憶容量大、汎化性低?”
　対数尤度を最大化するために、勾配法を用いることを考える.

$$
\begin{aligned}
\frac{\partial L(\theta)}{\partial \theta} &= -\sum_{i=1}^{N} \frac{\partial f_\theta(\mathbf{x}^{(i)})}{\partial \theta} - N \frac{\partial}{\partial \theta} \mathrm{log} Z(\theta) && (1.13) \\
&= -\sum_{i=1}^{N} \frac{\partial f_\theta(\mathbf{x}^{(i)})}{\partial \theta} - N \frac{1}{Z(\theta)} \int_{\mathbf{x}' \in X} -\exp(-f_\theta(\mathbf{x}')) \frac{\partial f_\theta(\mathbf{x}')}{\partial \theta} d\mathbf{x}' && (1.14) \\
&= -\sum_{i=1}^{N} \frac{\partial f_\theta(\mathbf{x}^{(i)})}{\partial \theta} + N \int_{\mathbf{x}' \in X} q_\theta(\mathbf{x}) \frac{\partial f_\theta(\mathbf{x}')}{\partial \theta} d\mathbf{x}' && (1.15) \\
&= -\sum_{i=1}^{N} \frac{\partial f_\theta(\mathbf{x}^{(i)})}{\partial \theta} + N \mathbb{E}_{\mathbf{x} \sim q_\theta(\mathbf{x})} \left[\frac{\partial f_\theta(\mathbf{x})}{\partial \theta} \right] && (1.16)
\end{aligned}
$$

ここで第二項は、生成モデルの確率分布に関して期待値を取ったものである。これをシュミレーションで求める際は、モンテカルロサンプリングを行うことが一般的であるが、計算コストが大きい.