

第2章 バイズ機械学習

2-1 確率統計

2-2 バイズ推論の基礎

2-3 確率的生成モデル

現実の問題では、データを生成する分布は複雑で1つの確率分布で取り扱えるケースは多くない。複数の分布をデータの生成過程を仮定しながら組み合わせて全体のモデル（同時分布）を作り、そこから事後分布、予測分布を計算する手法を「確率的生成モデル」と呼び、確率分布を複数組み合わせでできたモデルを「混合モデル」と呼ぶ。

2-3-1 混合モデルの構築

多峰性をもつデータに関してのクラスタリングを考える。データを表現するためのモデルを構築する要件定義として例えば以下の過程を考える。

1. K 個のクラスタは混合比率 $\pi = (\pi_1, \dots, \pi_K)$ で分布上に存在し、 π は事前分布 $p(\pi)$ から生成される。
2. それぞれのクラスタ自身の持つパラメータ θ_k が事前分布 $p(\theta_k)$ から生成される。
3. データ点 x_n が K 個ある分布うちのどれかから生成されるとし、 x_n に対応するクラスタの割り当てを s_n をする。この s_n は比率 π によって決まるとし、 s_n の生成する分布を $p(s_n|\vec{\pi})$ とする。
4. s_n によって選択された k 番目の確率分布 $p(\vec{x}_n|\vec{\theta}_k)$ からデータ x_n が生成される。

これら全ての確率分布をデータ生成順に組み合わせ、 N 個のデータに関して同時分布を考えると以下のようなになる。

$$p(X, S, \Theta, \vec{\pi}) = p(X|S, \Theta)p(S|\vec{\pi})p(\Theta)p(\vec{\pi}) = \left\{ \prod_{n=1}^N p(\vec{x}_n|\vec{s}_n, \Theta)p(\vec{s}_n|\vec{\pi}) \right\} \left\{ \prod_{k=1}^K p(\vec{\theta}_k) \right\} p(\vec{\pi})$$

実際に問題を解く際には、 $p(X|S, \Theta), p(\Theta)$ は問題設定に応じて決め、(クラスタリングの場合は) s_n をサンプリングする分布として以下のカテゴリ分布、

$$p(\vec{s}_n|\vec{\pi}) = \text{Cat}(\vec{s}_n|\vec{\pi}) = \prod_{k=1}^K \pi_k^{s_{n,k}}$$

π をサンプリングする分布としてカテゴリ分布の共役事前分布である *Dirichlet* 分布を選ぶことが多い。

$$p(\vec{\pi}) = \text{Dir}(\vec{\pi}|\vec{\alpha})$$

また s_n は直接は観測されないが、 x_n を生成する K 個の分布のうち 1 つを選択するという意味で、 x_n を発生させる確率分布を潜在的に決めている確率変数であると言える。このため s_n は潜在変数と呼ばれている。

2-3-2 混合モデルの推論

この同時分布から事後分布 $p(S, \Theta, \pi|X)$, クラスタ S の推定 $p(S|X)$ が可能であるが、いずれの計算も

$$p(X) = \sum_S \iint p(X, S, \Theta, \pi) d\Theta d\pi = \sum_s p(X, S)$$

$$p(S|X) = \iint p(S, \Theta, \pi|X) d\Theta d\pi$$

の計算が発生してしまい、解析的に解くことがほぼ不可能になる。次章で、この問題をある程度解消して近似的に解を出す方法を説明する。

2-4 近似推論

2-5 ガウス混合モデルと教師なし学習