

7章 最尤推定法における モデル選択

7.1 モデル選択とは

7.2 確率密度関数の近さを測る規準

7.1 モデル選択とは

さまざまな種類のガウスモデル

(A) 分散共分散行列 Σ が任意の正定値対称行列の場合

$$q(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

(B) 分散共分散行列 Σ が対角行列で、対角成分が異なる場合

$$q(x; \mu, \sigma^{(1)}, \dots, \sigma^{(d)}) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi(\sigma^{(j)})^2}} \exp\left(-\frac{(x^{(j)} - \mu^{(j)})^2}{2(\sigma^{(j)})^2}\right)$$

さまざまな種類のガウスモデル

(C)分散共分散行列 Σ が対角行列で、対角成分が等しい場合

$$q(x; \mu, \Sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{(x-\mu)^T(x-\mu)}{2\sigma^2}\right)$$

望ましいモデル

パラメトリック法で良い推定結果を得るために、
次のようなモデルを選ぶとよい

- ・ 真の確率密度関数をよく近似するモデル
- ・ パラメータ数の少ない単純なモデル

モデル選択

モデルの集合の中から、最も望ましいモデルを選択することをモデル選択という

どのモデルが適切かどうかは、真の確率分布と手元の訓練標本に依存して変化する

モデル選択

訓練標本によって最適なモデルを選ぶ場合の手順は以下のとおりである

- 1 いくつかのパラメトリックモデル $\{q_m(x; \theta)\}$ を用意する
- 2 各モデル m に対して、パラメータの最尤推定量 $\hat{\theta}_m$ を求める

モデル選択

- 3 各モデル m から得られた確率密度関数の推定量 $\hat{p}_m(x)$ について

$$\hat{p}_m(x) = q_m(x; \theta_m) \text{ で定める}$$

- 4 $\{\hat{p}_m(x)\}$ の中から、真の確率密度関数 $\hat{p}_m(x)$ に一番近いものを選び、それを最終的な確率密度関数の推定量とする

7.2 確率密度関数の近さを測る 規準

カルバック・ライブラー情報量

真の確率密度関数 $p(x)$ からの
カルバック・ライブラー情報量を評価尺度として
推定された確率密度関数 $\hat{p}(x)$ の良さを測る

カルバック・ライブラー情報量

カルバック・ライブラー情報量とは、真の確率密度関数 $p(x)$ と、任意の確率密度関数 $\hat{p}(x)$ の差の尺度である

$$KL(p \parallel \hat{p}) := \mathbb{E}_x \left[\log \frac{p(x)}{\hat{p}(x)} \right]$$

カルバック・ライブラー情報量

E_x は x に関する期待値を表す

$$E_x[f(x)] := \int f(x)p(x)dx$$

カルバック・ライブラー情報量

カルバック・ライブラー情報量は常に $KL(p \parallel \hat{p}) \geq 0$

$p(x) = \hat{p}(x)$ を満たすとき $KL(p \parallel \hat{p}) = 0$

カルバック・ライブラー情報量が小さいほど、
 $\hat{p}(x)$ は $p(x)$ の良い近似になっている

カルバック・ライブラー情報量

カルバック・ライブラー情報量の式を変形する

$$KL(p \parallel \hat{p}) = E_x [\log p(x)] - E_x [\log \hat{p}(x)]$$

$E_x [\log p(x)]$: 負のエントロピー (定数)

$E_x [\log \hat{p}(x)]$: 期待対数尤度

カルバック・ライブラー情報量

未知の確率密度関数 $p(x)$ が含まれていて直接計算できないので、 $p(x)$ に従うi.i.d標本 $\{x_i\}_{i=1}^n$ を用いて、平均対数尤度で近似する

$$E_x [\log \hat{p}(x)] \approx \frac{1}{n} \sum_{i=1}^n \log \hat{p}(x_i)$$

カルバック・ライブラー情報量

先述の平均対数尤度の式を最大にするモデルを選ぶことで、カルバック・ライブラー情報量が小さいモデルを選ぶことができる