第13章　モデル選択

13.1　導入

$X_1, X_2, \ldots, X_n$ を $\mathbb{X}$ 値の i.i.d. 確率変数列とし, Θ を母数空間とする. ただし, $\mathbb{X}$ と Θ は適当な距離空間の部分集合とする. 母数 $\Theta \ni \theta$ の「良さ」を測る観測に依存した関数である**コントラス** γ_n を考える. 具体的には

$$\gamma_n : \Theta \ni \theta \mapsto \gamma_n(\theta) := \gamma_n(X_1, X_2, \ldots, X_n, \theta) \in [0, \infty)$$

をコントラスト関数という. 多くの場合には, コントラスト関数は独立な確率変数の和

$$\gamma_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} c(X_j, \theta),$$

と書ける. ただし, $c(X_j \cdot) : \Theta \ni \theta \mapsto c(X_j, \theta) \in [0, \infty)$ である.

定義 13.1. 以下の用語を定義する.
(1) 母数 $\theta \in \Theta$ の**経験コスト**は $\gamma_n(\theta)$ である.
(2) **コスト**または**リスク**を $\mathsf{E}\big[\gamma_n(\theta)\big]$ で定める.
(3) $\Theta_m \subset \Theta$ を**統計的モデル**という. $\mathcal{M}$ を添え字集合とし, 統計的モデルの列を $\{\Theta_m\}_{m\in\mathcal{M}}$ と書いたとき, Θ_m を添え字 m の統計的モデルという.
(4) θ_* を目標母数といい

$$\theta_* \in \arg\min_{\theta\in\Theta} \mathsf{E}\big[\gamma_n(\theta)\big]$$

で定める. 特に, 統計的モデル $\mathcal{M}_m$ 上でのコスト関数の最小値

$$\theta_*^{(m)} \in \arg\min_{\theta\in\Theta_m} \mathsf{E}\big[\gamma_n(\theta)\big]$$

を**射影された目標母数**という.
(5) 各モデル Θ_m に対して, M **推定量**を

$$\widehat{theta}^{(m)} \in \arg\min_{\theta\in\Theta_m} \gamma_n(\theta)$$

で定める.

(6) モデル列 $\{\Theta_m\}_{m\in\mathcal{M}}$ の中で**最適なモデル** Θ_{m_*} をその M 推定量のコストを最小にするもので定義する. すなわち

$$m_* \in \arg\min_{m\in\mathcal{M}} \mathsf{E}\big[\gamma_n(\widehat{\theta}^{(m)})\big]$$

である.

$\boxed{\text{問題}}$ 統計的モデルの列 $\{\Theta_m\}_{m\in\mathcal{M}}$ が与えれたとき, 統計的モデル Θ_m と Θ_m の要素である推定量 $\widehat{\theta}$ をうまく選ぶことができるか? また, θ_* のよい推定量となっているか?

$\boxed{\text{過学習}}$

説明を加える.

13.2 罰則化

定義 13.2. 統計的モデルの族 $\{\Theta\}_{m\in\mathcal{M}}$ に対する**罰則関数**を

$$\mathsf{pen} : \mathcal{M} \longrightarrow [0, \infty)$$

で定める. なお, 罰則関数 pen は観測 $X_1, X_2, \ldots, X_n$ に依存してもよい.

新たな推定量 $\widehat{m}$ を

$$\widehat{m} \in \arg\min_{m\in\mathcal{M}} \big\{\gamma_n(\widehat{\theta}^{(m)}) + \mathsf{pen}(m)\big\} \tag{13.1}$$

で定める. 以後では, (13.1) で定まる推定量 $\widehat{\theta}^{\widehat{m}}$ を $\widetilde{\theta}$ と記すことにする.