Regression with stagewise minimization on risk function

鹿児島大・理工 吉田拓真，　島根大・総合理工 内藤貫太

本講演では, 逐次推定に基づくノンパラメトリック回帰推定法を提案する．確率変数 $(Y, X) \in (\mathcal{Y}, \mathcal{X}) \subseteq (\mathbb{R}, \mathbb{R}^p)$ に対して, $\mu(x) = E[Y|X = x]$ とその近似関数 μ_0 のリスク関数を

$$D_U(\mu, \mu_0) = E[U(\xi(\mu_0(X))) - U(\xi(\mu(X))) - Y\{\xi(\mu_0(X)) - \xi(\mu(X))\}]$$

と定義する． ただし, U は凸関数で, $\xi = (U')^{-1}$ である． 例えば, $U(x) = 2^{-1}x^2$ とすると, $D_U(\mu, \mu_0) = 2^{-1}E[(Y - \mu_0(X))^2] - 2^{-1}E[Y^2]$ となり, 最小二乗法に帰着する． このように, U の選択に応じて様々なリスク関数を表現することができる．特に, $D_U(\mu, \mu_0)$ は指数型分布族のカルバック・ライブラー情報量を含むことが簡単に示せる．経験リスク (経験分布関数によって近似されたリスク関数) を最小とする μ_0 を μ の推定量として構成する．まず, 複数のワードの集合である辞書 $\mathcal{D} = \{B_k(x)|k = 1, \cdots, K, K < \infty\}$ を用意する．このとき, μ の推定量を以下の形で定義する:

$$\hat{\mu}(x) = U'\left(\sum_{t=1}^{M} \lambda_t \xi(B_t(x))\right), \quad B_1, \cdots, B_M \in \mathcal{D}.$$

ここで, $\lambda_1, \cdots, \lambda_M$ は重み係数で, $\lambda_t \geq 0, \sum_{t=1}^{M} \lambda_t = 1$ を満たすものとする．つまり, 推定量は ξ で変換されたワードの凸結合で表されるとする．そこで我々は経験リスク最小化に基づく逐次アルゴリズムを提案し, 各ワード B_t は各ステップで辞書の中からひとつずつ選択する．したがって, 推定量 $\hat{\mu}(x)$ に用いるワードの個数 M とはアルゴリズムの更新回数を意味する．提案するアルゴリズムでは, M 回の反復更新の中でひとつのワードが複数回選択されることや, 辞書の中で全く選択されないワードが存在することもある．このような特徴から, 本手法はブースティングに近いアルゴリズムであるものの, 提案手法が異なる点は, 重み係数 λ_t が最適化されるものではなく, アルゴリズムで用いられる混合係数から自然に導かれる点である．そのため, 提案手法はブースティングよりもシンプルなアルゴリズムであると言える．

提案した推定量の理論的性質として, リスク関数の期待値 $E[D_U(\mu, \hat{\mu})]$ の非漸近的誤差限界を導出する．この証明には, 推定量が提案したアルゴリズムを用いて構成されることを用いる．したがって, 各ワードが逐次的に選択されること, 推定量を凸結合で構成すること, 重み係数はデータに依存せずに決定される (つまり, これらは確率変数ではない) ことが本質的に重要となっている．結果から, アルゴリズムの更新回数 M を増やすと $E[D_U(\mu, \hat{\mu})]$ の非漸近的誤差限界が減少することがわかった．このようなアプローチ, 理論結果は Klemelä (2007) と Naito and Eguch (2013) が密度推定において提案している．したがって, 本研究の位置づけとしては, 上記研究の回帰分析における展開と捉えることができる．

本講演では, 提案手法に関するリスク関数, 逐次アルゴリズム, 非漸近的誤差限界，およびシミュレーションによる数値例の結果を報告する．特に数値例では, 本研究の重要な応用問題として $\mu(x)$ がスパース加法モデルで定義されているケースを扱い, 提案手法が高次元ノンパラメトリック回帰における罰則付き最適化法とは異なる新たなアプローチとなることを示す．

References

Klemelä,J. (2007). Density estimation with stagewise optimization of the empirical risk. *Machine Learning.* **67** 169–195.

Naito,K. and Eguchi,S. (2013). Density estimation with minimization of U-divergence. *Machine Learning.* **90** 29–57.