

Data Science and Machine Learning

照屋 佑喜仁

June 10, 2025

1 2.4 Tradeoffs in Statical Learning

- 教師あり学習の技術
- Tradeoff
- irreducible risk, approximation error, statistical error
- approximation-estimation tradeoff

教師あり学習

- 教師あり学習における機械学習の技術
 - generalization risk(2.5) あるいは expected generalization risk(2.6) をできるだけ小さくする
 - できるだけ少ない計算リソースで
- これを達成するために、適切な予測関数の集合 \mathcal{G} を選ぶ必要がある。この選び方は下のような要因によって決まる。
 - 集合の複雑さ (最適な予測関数 g^* を適切に近似, あるいは含むのに十分に複雑 (豊か) か?)
 - (2.4) の最適化によって学習者を訓練する容易さ
 - 集合 \mathcal{G} において, training loss(2.3) が risk(2.1) をどれだけ正確に推定するか
 - 連続なのか, 分類なのか……

Tradeoff

- 集合 \mathcal{G} の選択は、通常トレードオフを伴う
 - 単純な \mathcal{G} からの学習器は早く訓練できるが、上手く近似できない可能性
 - g^* を含むような豊かな \mathcal{G} からの学習器は多くの計算リソースを必要とする可能性
- モデルの複雑さ、計算の単純さ、推定の制度の関係を見るために2つの tradeoff について考えていく
 - the approximation-estimation tradeoff(近似-推定トレードオフ)
 - the bias-variance tradeoff(バイアス-分散トレードオフ)

今, generalization risk(2.5) を3つの要素に分解して考える.

$$\ell(g_{\tau}^{\mathcal{G}}) = \underbrace{\ell^*}_{\text{irreducible risk}} + \underbrace{\ell(g^{\mathcal{G}}) - \ell^*}_{\text{approximation error}} + \underbrace{\ell(g_{\tau}^{\mathcal{G}}) - \ell(g^{\mathcal{G}})}_{\text{statistical error}} \quad (2.16)$$

irreducible risk, approximation error

$$\ell(g_\tau^{\mathcal{G}}) = \underbrace{\ell^*}_{\text{irreducible risk}} + \underbrace{\ell(g^{\mathcal{G}}) - \ell^*}_{\text{approximation error}} + \underbrace{\ell(g_\tau^{\mathcal{G}}) - \ell(g^{\mathcal{G}})}_{\text{statistical error}} \quad (2.16)$$

- ℓ^* は $\ell(g^*)$ で定義される irreducible risk(還元不能リスク). どの学習器も ℓ^* より小さいリスクで予測することはできない.
- $g^{\mathcal{G}}$ は $\operatorname{argmin}_{g \in \mathcal{G}} \ell(g)$ で定義される, \mathcal{G} 内で最も最良の学習器.
- $\ell(g^{\mathcal{G}}) - \ell^*$ は approximation error(近似誤差). irreducible risk と \mathcal{G} のなかで最良の予測関数の risk の差を見ている.
 - 適切な \mathcal{G} を選び, その上で $\ell(g)$ を最小化するのは, 単純に数値解析と関数解析の問題となる (ここで訓練データ τ は登場しないから)
 - \mathcal{G} が g^* を含まなければ近似誤差は任意に小さく出来ず risk を大きくする要因となる
 - 近似誤差を減らす唯一の方法は, \mathcal{G} を大きくしてより多くの関数を含めること

statical (estimation) error

$$\ell(g_\tau^{\mathcal{G}}) = \underbrace{\ell^*}_{\text{irreducible risk}} + \underbrace{\ell(g^{\mathcal{G}}) - \ell^*}_{\text{approximation error}} + \underbrace{\ell(g_\tau^{\mathcal{G}}) - \ell(g^{\mathcal{G}})}_{\text{statistical error}} \quad (2.16)$$

- $\ell(g_\tau^{\mathcal{G}}) - \ell(g^{\mathcal{G}})$ は statistical(estimation) error(統計的 (推定) 誤差). 訓練セット τ に依存. 特に, 学習器 $g_\tau^{\mathcal{G}}$ が \mathcal{G} の最良の予測関数 $g^{\mathcal{G}}$ をどれだけ上手く推定しているかに依存している.
 (良い予測器なら) この誤差は訓練サイズが無限大に近づくにつれて(確率的に, または期待値として)0 に収束するはずである.

approximation-estimation tradeoff

approximation-estimation tradeoff(近似-推定トレードオフ)は、2つの相反する要求を対立させる。

- \mathcal{G} が十分にシンプルで、統計的誤差が大きくなりすぎない必要がある。(推定しやすい?)
- \mathcal{G} が十分に充実して、近似誤差が小さいことを保証する必要がある。(g^* をできれば見つけたい?)

2乗誤差損失でのリスクを解釈してみる

2乗誤差損失のリスクは $\ell(g_\tau^G) = \mathbb{E}[(Y - g_\tau^G(\mathbf{X}))^2]$ となる。このとき、最適な予測関数は $g^*(\mathbf{x}) = \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$ で与えられるのであった。
(theorem2.1)

このとき、分解 (2.16) は以下のように解釈できる。

- $\ell^* = \mathbb{E}[(Y - g^*(\mathbf{X}))^2]$ は還元不能誤差であり、これより小さい期待2乗誤差の予測関数はない。

-