

問題A

各ハイパーパラメータの重要度を評価する既存手法として、**functional analysis of variance (functional ANOVA)** と呼ばれる手法がある。

あるアルゴリズム(機械学習モデル) A が n 個のハイパーパラメータを持っているとする。また、 $1 \leq i \leq n$ で表される(インデックス付けされる)ハイパーパラメータが取りうる値の集合を Θ_i とおく。つまり、ハイパーパラメータが全体として取りうる値の集合は、 $\Theta = \prod_{i=1}^n \Theta_i$ である。 $N = \{1, \dots, n\}$ としたとき、 $U \subseteq N$ (ここで、 U が空でないときは、 $U = \{i_1, \dots, i_m\}$, $1 \leq i_1 < \dots < i_m \leq n$ とおく。)で指定できるハイパーパラメータを特定の値に固定したものを $\theta_U = \langle \theta_{i_1}, \dots, \theta_{i_m} \rangle$ と書く。このとき、特に $\theta_N \in \Theta$ である。

functional ANOVAは、任意の $\theta \in \Theta$ に対して予測される A の性能 $\hat{y}(\theta) \in \mathbb{R}$ を返す再帰モデル(random forest)の上に作られている。そこで、 $\hat{a}_U(\theta_U) := (\theta'_U = \theta_U$ となるような $\hat{y}(\theta')$ の平均値)が定義できる。この $\hat{a}_U(\theta_U)$ の定義における項 $\hat{y}(\theta')$ の数は $N \setminus U$ で指定できるハイパーパラメータたちの自由度に依るので指数的な大きさを持つと言えるが、 $\hat{a}_U(\theta_U)$ はrandom forestの葉(末端ノード)の数に対して線形に計算できる。

functional ANOVAは上で述べた $\hat{y} : \prod_{i=1}^n \Theta_i \rightarrow \mathbb{R}$ を以下を満たす関数族 $\{f_U\}_{U \subseteq N}$ に分解する。

$$\bullet \quad \hat{y}(\theta) = \sum_{U \subseteq N} \hat{f}_U(\theta_U)$$

$$\bullet \quad \hat{f}_U(\theta_U) = \begin{cases} \hat{f}_\emptyset & (U = \emptyset) \\ \hat{a}_U(\theta_U) - \sum_{W \subsetneq U} \hat{f}_W(\theta_W) & (U \neq \emptyset) \end{cases}$$

(ここで \hat{f}_\emptyset は、 Θ 全体での $\hat{y}(\theta)$ の平均値を値としてとる定数関数とする。)

このとき、 $\hat{f}_{\{j\}}(\theta_{\{j\}})$ がハイパーパラメータ j ($1 \leq j \leq n$)の \hat{y} への寄与を表していることが分かる。各ハイパーパラメータが \hat{y} の分散 V にどの程度影響を及ぼしているかをその「ハイパーパラメータの重要度」だと思ふことにすると、ハイパーパラメータの重要度は V_j/V と思ふことができる。(ここで、 $V_j = \frac{1}{\|\Theta_j\|^2} \int \hat{f}_{\{j\}}(\theta_{\{j\}})^2 d\theta_{\{j\}}$ としている。)

理論面ないし実用面での制限

functional ANOVA自体は一つのデータセットさえあれば実行できるアルゴリズムであるが、そのデータセット上で選定されるハイパーパラメータが必ずしも他のデータセットにおいても一般的に重要であるとは限らない点に注意が必要である。また、例えば、 \hat{y} がほぼ定数関数になってしまうようなデータセットの上では、どれが重要なハイパーパラメータなのかを選定することは困難である。

上で述べた制限への対処法

functional ANOVAを用いて重要なハイパーパラメータを選定する場合、多くのデータセット上で functional ANOVAを適用するのが望ましい。

参考文献

1. [Hyperparameter Importance Across Datasets](#)
Jan N. van Rijn, Frank Hutter 2018