

# [問題A]

各ハイパーパラメータの重要度を評価するための既存手法として、functional ANOVA というものがある。

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。また、 $1 \leq i \leq n$  で表される (1-次元入力付きである)

ハイパーパラメータ  $i$  に対応する値の集合を  $\Theta_i$  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、 $U \subseteq N$  に含まれるハイパーパラメータを特定の値に固定したものを  $\Theta_U = \langle \theta_1, \dots, \theta_{\dim} \rangle$  ( $U = \{i_1, \dots, i_m\}$ ) と書く。このとき、 $\Theta_U \in \Theta$  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  $y(\theta) \in \mathbb{R}$  を変数再帰モデルの上で作られている。

そこで  $\hat{y}_U(\theta_U) := (\theta'_U = \theta_U$  となるような  $y(\theta')$  の平均値) が定義できる。この  $\hat{y}_U(\theta_U)$  における項  $y(\theta')$  の数は  $N \setminus U$  に含まれるパラメータの自由度によるに指数的な大きさを持つ数であるが、functional ANOVA における項の数に対して線形に計算ができる。

functional ANOVA は上で述べた  $y: \Theta_1 \times \dots \times \Theta_n \rightarrow \mathbb{R}$  を以下で与えられる関数  $\{f_U\}$  に分解する。

$$y(\theta) = \sum_{U \subseteq N} f_U(\theta_U)$$

$$\hat{f}_U(\theta_U) = \begin{cases} \hat{f}_\emptyset = (\Theta \text{ 全体での } y(\theta) \text{ の平均値}) & (U = \emptyset \text{ のとき}) \\ \hat{y}_U(\theta_U) - \sum_{w \subseteq U} \hat{f}_w(\theta_w) & (U \neq \emptyset \text{ のとき}) \end{cases}$$

このとき、 $\hat{f}_{\{i\}}(\theta_{\{i\}})$  がパラメータ  $i$  ( $1 \leq i \leq n$ ) の  $y$  への寄与を表していることが分かる。各パラメータが  $y$  の分散  $V$  に対してどの程度影響しているかを「パラメータの重要度」と思うことにすると、パラメータの重要度は  $V_i/V$  と考えることができる。

$$\text{そこで、} V_i = \frac{1}{\|\Theta_{\{i\}}\|^2} \int \hat{f}_{\{i\}}(\theta_{\{i\}})^2 d\theta_{\{i\}} \text{ としている。}$$

## [問題A] 続き

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上では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、 $\gamma$  がほぼ定数関数になる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の上で 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  
arXiv:1710.04725v2