

A PROOF

Lemma 1. The binary relation $\preceq_{(Z)}$ defined in Eq. (7) is transitive, i.e., $\forall \mathbf{x}, \mathbf{y}, \mathbf{w} \in X$, $F(\mathbf{x}) \preceq_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \preceq_{(Z)} F(\mathbf{w}) \Rightarrow F(\mathbf{x}) \preceq_{(Z)} F(\mathbf{w})$.

Proof of Lemma 1. According to Eq. (7),

$$\begin{aligned}
 & F(\mathbf{x}) \preceq_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \preceq_{(Z)} F(\mathbf{w}) \Rightarrow \\
 & (F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \vee F(\mathbf{x}) =_{(Z)} F(\mathbf{y})) \wedge (F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}) \vee F(\mathbf{y}) =_{(Z)} F(\mathbf{w})) \Rightarrow \\
 & \underbrace{(F(\mathbf{x}) =_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) =_{(Z)} F(\mathbf{w}))}_i \vee \underbrace{(F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}))}_{ii} \\
 & \vee \underbrace{(F(\mathbf{x}) =_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}))}_{iii} \vee \underbrace{(F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) =_{(Z)} F(\mathbf{w}))}_{iv}
 \end{aligned} \tag{9}$$

Statement I: $F(\mathbf{x}) =_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) =_{(Z)} F(\mathbf{w}) \Rightarrow F(\mathbf{x}) =_{(Z)} F(\mathbf{w})$.

Proof of Statement I. From Eq. (5), we have,

$$\begin{aligned}
 & F(\mathbf{x}) =_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) =_{(Z)} F(\mathbf{w}) \Rightarrow \\
 & (f^k(\mathbf{x}) = f^k(\mathbf{y}) \vee (f^k(\mathbf{x}) \leq z^k \wedge f^k(\mathbf{y}) \leq z^k), \forall k \in [K]) \\
 & \wedge (f^k(\mathbf{y}) = f^k(\mathbf{w}) \vee (f^k(\mathbf{y}) \leq z^k \wedge f^k(\mathbf{w}) \leq z^k), \forall k \in [K]) \\
 & \Rightarrow \forall k \in [K] : (f^k(\mathbf{x}) = f^k(\mathbf{y}) = f^k(\mathbf{w})) \\
 & \vee ((f^k(\mathbf{x}) = f^k(\mathbf{y})) \wedge (f^k(\mathbf{y}) \leq z^k \wedge f^k(\mathbf{w}) \leq z^k)) \\
 & \vee ((f^k(\mathbf{x}) \leq z^k \wedge f^k(\mathbf{y}) \leq z^k) \wedge (f^k(\mathbf{y}) = f^k(\mathbf{w}))) \\
 & \vee (f^k(\mathbf{x}) \leq z^k \wedge f^k(\mathbf{y}) \leq z^k \wedge f^k(\mathbf{w}) \leq z^k) \\
 & \Rightarrow ((f^k(\mathbf{x}) = f^k(\mathbf{w})) \vee (f^k(\mathbf{x}) \leq z^k \wedge f^k(\mathbf{w}) \leq z^k), \forall k \in [K]) \\
 & \Rightarrow F(\mathbf{x}) =_{(Z)} F(\mathbf{w})
 \end{aligned} \tag{10}$$

Statement II: $F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}) \Rightarrow F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w})$

Proof of Statement II. According to Eq. (5), we have

$$\begin{aligned}
 & F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \Leftrightarrow \\
 & \exists k_1 \in [K] : f^{k_1}(\mathbf{x}) < f^{k_1}(\mathbf{y}) \wedge f^{k_1}(\mathbf{y}) > z^{k_1} \wedge F^{k_1-1}(\mathbf{x}) =_{(Z)} F^{k_1-1}(\mathbf{y}) \\
 & F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}) \Leftrightarrow \\
 & \exists k_2 \in [K] : f^{k_2}(\mathbf{y}) < f^{k_2}(\mathbf{w}) \wedge f^{k_2}(\mathbf{w}) > z^{k_2} \wedge F^{k_2-1}(\mathbf{y}) =_{(Z)} F^{k_2-1}(\mathbf{w})
 \end{aligned} \tag{11}$$

Let $k' = \min\{k_1, k_2\}$, according to Statement I and Eq. (11), we have:

$$(F^{k'-1}(\mathbf{x}) =_{(Z)} F^{k'-1}(\mathbf{y})) \wedge (F^{k'-1}(\mathbf{y}) =_{(Z)} F^{k'-1}(\mathbf{w})) \Rightarrow F^{k'-1}(\mathbf{x}) =_{(Z)} F^{k'-1}(\mathbf{w}) \tag{12}$$

According to Eq. (12), if $k' = k_1 < k_2$, we have

$$(f^{k'}(\mathbf{x}) < f^{k'}(\mathbf{y}) = f^{k'}(\mathbf{w})) \wedge (z^{k'} < f^{k'}(\mathbf{y}) = f^{k'}(\mathbf{w})) \tag{13}$$

If $k' = k_2 < k_1$, we have

$$(f^{k'}(\mathbf{x}) = f^{k'}(\mathbf{y}) < f^{k'}(\mathbf{w})) \wedge (z^{k'} < f^{k'}(\mathbf{w})) \tag{14}$$

Otherwise when $k' = k_1 = k_2$, we have

$$\left(f^{k'}(\mathbf{x}) < f^{k'}(\mathbf{y}) < f^{k'}(\mathbf{w})\right) \wedge \left(z^{k'} < f^{k'}(\mathbf{y}) < f^{k'}(\mathbf{w})\right) \quad (15)$$

Combing Eq. (12), Eq. (13), Eq. (14) and Eq. (15), we have found $k = k'$ such that $(f^k(\mathbf{x}) < f^k(\mathbf{w})) \wedge (f^k(\mathbf{w}) > z^k) \wedge (F^{k-1}(\mathbf{x}) =_{(Z)} F^{k-1}(\mathbf{w}))$.

Statement III: $F(\mathbf{x}) =_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w}) \Rightarrow F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w})$.

Proof of Statement III. According to the definitions of $=_{(Z)}$ and $\prec_{(Z)}$ in Eq. (5) and Eq. (6),

$$\begin{aligned} & \left(F(\mathbf{x}) =_{(Z)} F(\mathbf{y})\right) \wedge \left(F(\mathbf{y}) \prec_{(Z)} F(\mathbf{w})\right) \Rightarrow \\ & \left(\forall k_1 \in [K] : (f^{k_1}(\mathbf{x}) = f^{k_1}(\mathbf{y})) \vee (f^{k_1}(\mathbf{x}) \leq z^{k_1} \wedge f^{k_1}(\mathbf{y}) \leq z^{k_1})\right) \\ & \wedge \left(\exists k_2 \in [K] : f^{k_2}(\mathbf{y}) < f^{k_2}(\mathbf{w}) \wedge f^{k_2}(\mathbf{w}) > z^{k_2} \wedge F^{k_2-1}(\mathbf{y}) =_{(Z)} F^{k_2-1}(\mathbf{w})\right) \end{aligned} \quad (16)$$

According to the above propositions, we find a $k = k_2$, such that,

$$\begin{aligned} & \left((f^k(\mathbf{x}) = f^k(\mathbf{y}) < f^k(\mathbf{w})) \vee (f^k(\mathbf{x}) \leq z^k < f^k(\mathbf{w}))\right) \\ & \wedge \left(z^k < f^k(\mathbf{w})\right) \wedge \left(F^{k-1}(\mathbf{x}) =_{(Z)} F^{k-1}(\mathbf{w})\right) \\ & \Rightarrow \left(f^k(\mathbf{x}) < f^k(\mathbf{w})\right) \wedge \left(z^k < f^k(\mathbf{w})\right) \wedge \left(F^{k-1}(\mathbf{x}) =_{(Z)} F^{k-1}(\mathbf{w})\right) \end{aligned} \quad (17)$$

in which indicates $F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w})$.

Statement IV: $F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) =_{(Z)} F(\mathbf{w}) \Rightarrow F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w})$.

Proof of Statement IV. According to the definitions of $=_{(Z)}$ and $\prec_{(Z)}$ in Eq. (5) and Eq. (6),

$$\begin{aligned} & \left(F(\mathbf{x}) \prec_{(Z)} F(\mathbf{y})\right) \wedge \left(F(\mathbf{y}) =_{(Z)} F(\mathbf{w})\right) \\ & \Rightarrow \left(\exists k_1 \in [K] : f^{k_1}(\mathbf{x}) < f^{k_1}(\mathbf{y}) \wedge f^{k_1}(\mathbf{y}) > z^{k_1} \wedge F^{k_1-1}(\mathbf{x}) =_{(Z)} F^{k_1-1}(\mathbf{y})\right) \wedge \\ & \left(\forall k_2 \in [K] : (f^{k_2}(\mathbf{y}) = f^{k_2}(\mathbf{w})) \vee (f^{k_2}(\mathbf{y}) \leq z^{k_2} \wedge f^{k_2}(\mathbf{w}) \leq z^{k_2})\right) \end{aligned} \quad (18)$$

Then, we can find a $k = k_1$, such that,

$$\begin{aligned} & \left(f^k(\mathbf{x}) < f^k(\mathbf{y}) = f^k(\mathbf{w})\right) \wedge \left(z^k < f^k(\mathbf{w}) = f^k(\mathbf{y})\right) \wedge \left(F^{k-1}(\mathbf{x}) =_{(Z)} F^{k-1}(\mathbf{w})\right) \\ & \Rightarrow \left(f^k(\mathbf{x}) < f^k(\mathbf{w})\right) \wedge \left(z^k < f^k(\mathbf{w})\right) \wedge \left(F^{k-1}(\mathbf{x}) =_{(Z)} F^{k-1}(\mathbf{w})\right) \Rightarrow F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w}) \end{aligned} \quad (19)$$

By substituting the conclusions in statement (I), (II), (III) and (IV) into (i), (ii), (iii), and (iv) in Eq. (9) respectively, we get:

$$\begin{aligned} F(\mathbf{x}) \preceq_{(Z)} F(\mathbf{y}) \wedge F(\mathbf{y}) \preceq_{(Z)} F(\mathbf{w}) & \Rightarrow \left(F(\mathbf{x}) \prec_{(Z)} F(\mathbf{w})\right) \vee \left(F(\mathbf{x}) =_{(Z)} F(\mathbf{w})\right) \\ & \Rightarrow F(\mathbf{x}) \preceq_{(Z)} F(\mathbf{w}) \end{aligned} \quad (20)$$

which concludes the proof. \square

B SUPPLEMENTARY RESULTS

B.1 FEATURE SELECTION IN BIOINFORMATICS

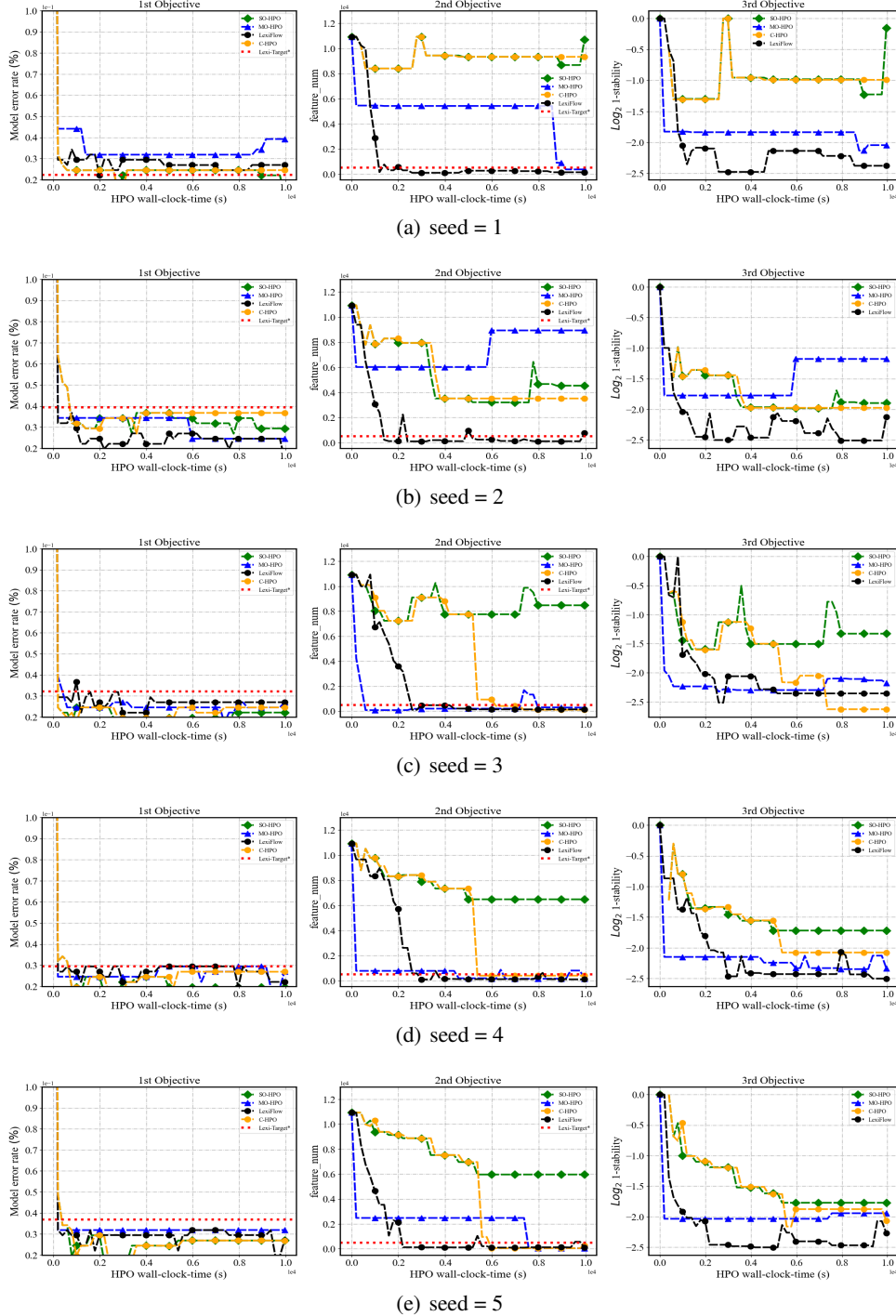


Figure 4: The detailed results in tuning Xgboost on biological dataset AP_Colon_Kidney

C EXPERIMENTATION DETAILS

C.1 DETAILED SEARCH SPACE

The detailed search space of the hyperparameters in tuning Neural Networks, Xgboost and Random Forest.

Table 3: Hyperparameters tuned in Random Forest

hyperparameter	type	range
max features	float	$[\min(0.1, 1/\sqrt{\text{data_features}}), 1.0]$
estimators number	int	$[4, \min(2048, \text{train_datasize})]$
max leaves	int	$[4, \text{train_size}]$

Table 4: Hyperparameters tuned in XGboost

hyperparameter	type	range
estimators number	int	$[4, \min(32768, \text{train_datasize})]$
max leaves	int	$[4, \min(32768, \text{train_datasize})]$
max depth	int	$[0, 6, 12]$
min child weight	float	$[0.001, 128]$
learning rate	float	$[1/1024, 1.0]$
subsample	float	$[0.1, 1.0]$
colsample by tree	float	$[0.01, 1.0]$
colsample by level	float	$[0.01, 1.0]$
reg alpha	float	$[1/1024, 1024]$
reg lambda	float	$[1/1024, 1024]$

Table 5: Hyperparameters tuned in Neural Network

hyperparameter	type	range
epoch num	int	$[1, 20]$
layer num	int	$[1, 3]$
hidden units num per layer	int	$[4, 128]$
dropout value per layer	float	$[0.2, 0.5]$
learning rate per layer	float	$[1e-5, 1e-1]$

C.2 DETAILED DATASETS INFORMATIONS IN TABLE 2

Table 6: Date statistics information

Dataset statistics	Gisette	Christine	Scene	Ionsphere	Ginal_prior	Ginal_agnostic	Bioresponse	Hiva_agnostic	Madelon
# of train instance	4900	4063	1805	263	2601	2601	2813	3171	1950
# of val instance	2098	1355	602	88	867	867	938	1058	650
# full feature dimension	5000	1636	299	34	784	970	1776	1617	500

Are datasets used in Section 4.2 are available in Openml <https://www.openml.org/>.