Technical Report NDF-SR

Anonymous ICDM author

I. Capability Study of NDF-SR

In this section, we give a full study of the capability of the NDF-SR model, including the high capability of NDF-SR and how the NDT-pruning control the capability.

For a clear mathematical representation, we use *Degrees of Freedom* (DoF) to formally define the capability of a model. The DoF is defined as follows: Assuming that we have data of the form $D_n = \{(\boldsymbol{X}_1, Y_1), ..., (\boldsymbol{X}_n, Y_n)\}$ where each $\boldsymbol{X}_i \in \mathbb{R}^{n'}$ denotes a vector of n' features, $Y_i \in \mathbb{R}$ denotes the response. The relationship of the variables is in the form of:

$$Y_i = f(\boldsymbol{X}_i) + \epsilon_i \tag{1}$$

where f is the underlying truth function; the errors $\epsilon_1, ..., \epsilon_n$ are uncorrelated and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)(\forall i)$. If our model gives an predicted relation between $\boldsymbol{X}$ & Y as $\hat{f}$ (i.e., $\hat{Y}_i = \hat{f}(\boldsymbol{X}_i)$), then, the DoF is defined as:

$$DoF(\hat{f}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \text{Cov}(\hat{Y}_i, Y_i) \tag{2}$$

Since in real-world tests, we do not know the underlying true function f. Thus, we perform the analysis of DoF on two simulated underlying functions. The first one is "MARSadd" underlying f proposed by [1]:

$$Y = 0.1e^{4X_1} + \frac{4}{1 + e^{-20(X_2 - 0.5)}} + 3X_3 + 2X_4 + X_5 + \epsilon \tag{3}$$

In the experiments, we take $X_i \sim Unif(0, 1),\ \epsilon \sim \mathcal{N}(0, 1)$, where $Unif$ means uniform distribution. This underlying true function aims to test the DoF property under complex exponential behavior.

The second underlying f we proposed is:

$$Y' = \sum_{i=1}^{5} X_i^4 + 0 \cdot \sum_{i=6}^{10} X_i^4 + \epsilon \tag{4}$$

It is denoted as "POWERadd" and aims to test the model's DoF behavior under the complex curve and extra dimension. For the experiments based on this underlying function, we take X_i and ϵ independently from the standard normal distribution. [1]

The results of DoF on simulated data are shown in Fig.1. We can see that no matter under which setting, except for some extreme cases (like $r = 0.9$ or $depth = 1$), the NDF-SR generally has a significantly higher DoF than the linear model, which proves the target for using NDF-SR to provide DoF that the linear model lacks. Also, as the depth of the NDT increase, the model has a higher DoF. Furthermore, in both experiments, we can see that the DoF decreases as we increase the pruning rate. This supports that the NDT-pruning fixes the overfitting problem by controlling DoF to make it fit the problem. To conclude, this section shows the effectiveness of our NDF-SR model in giving reasonable capability to the model. More importantly, under the most extreme case (i.e., $depth = 1$), the DoF of NDF-SR is still generally higher than the linear model. This shows the superiority in fixing the lacking DoF of the NDF-SR enhancement.

[1]Note that we choose a different initialization for $\boldsymbol{X}$ in MARSadd and POWERadd, this is because 1 states that normally we should use the standard normal to initialize, but for MARSadd, we use uniform distribution to initialize.

References

[1] L. Mentch and S. Zhou, "Randomization as regularization: A degrees of freedom explanation for random forest success," 2020.

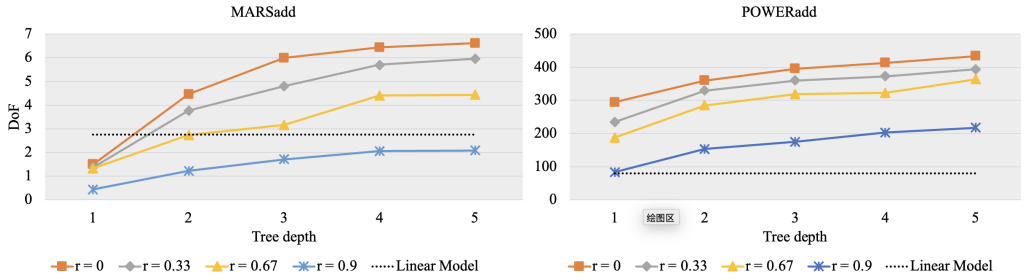

Fig. 1: Degrees of freedom for NDF-SR with different depths and different pruning rates (r) on simulated data 'MARSadd' model and 'POWERadd' model. Every point represents a Monte Carlo approximation of DoF for 200 trials.