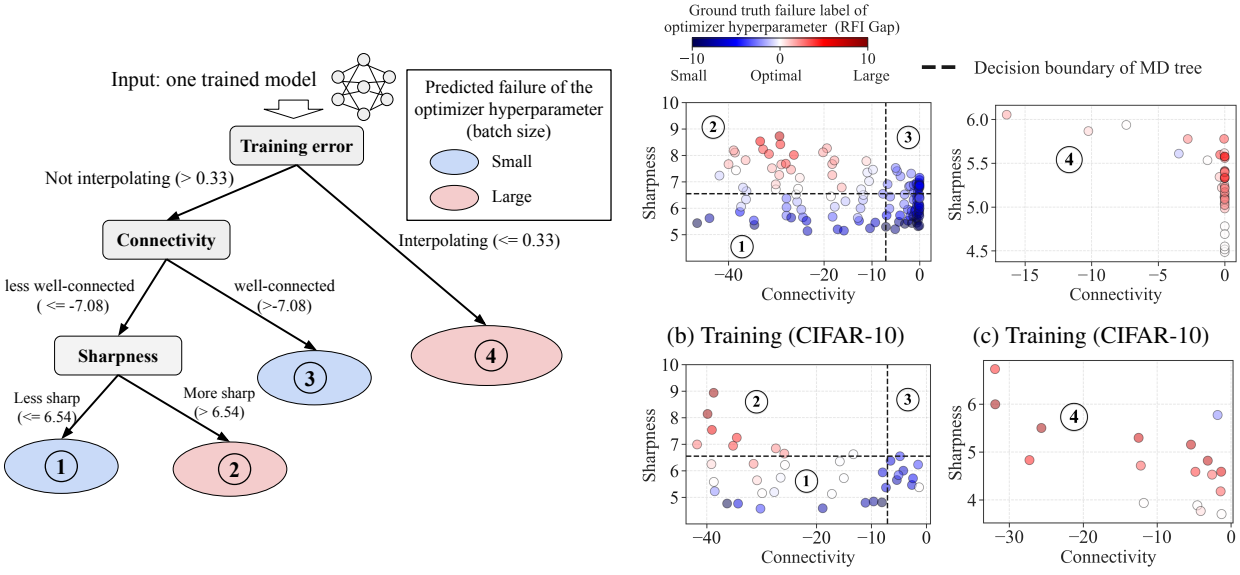
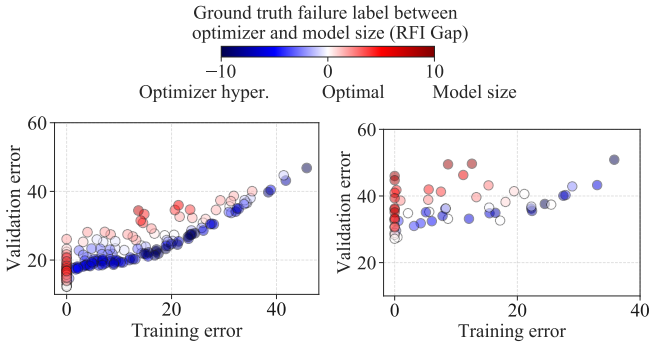


(a) MD tree for Q1 (b) Training (CIFAR-10) (c) Training (CIFAR-10) (d) Test (CIFAR-10-LT) (e) Test (CIFAR-10-LT)

Figure 26. **(Visualizing MD tree and its diagnosis results on models trained on imbalanced data (CIFAR-10-LT)).** *Left side*: structure of the MD tree. The color of the leaf node indicates the predicted class by `MD tree`. The threshold values are learned from the training set. *Right side*: The first row represents training samples, and the second row represents test samples. Each colored circle represents one sample (which is one pre-trained model), and the color represents the ground-truth label: blue means the hyperparameter is too large, while red means small. The black dashed line indicates the decision boundary of `MD tree`. Each numbered regime on the right corresponds to the leaf node with the same number on the tree. The samples in 26(b) and those in 26(c) are separated by training error. The same applies to 26(d) and 26(e).



(a) Training (CIFAR-10) (b) Test (CIFAR-10-LT)

Figure 27. **(Applying validation metrics to diagnose the models trained with imbalanced data (CIFAR-10-LT).)** (a)training set comprises models trained with CIFAR-10, test set comprises models trained with CIFAR-10-LT.

Figure 26 shows that the decision boundary of loss landscape metrics learned by the MD tree from CIFAR-10 models can well transfer to CIFAR-10-LT models. Figure 27 shows that there exists large distribution shift of validation metrics between CIFAR-10 and CIFAR-10-LT models.