

Figure 4. (Visualizing MD tree and its diagnosis results for Q1). *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 4(b) and those in 4(c) are separated by training error. The same applies to 4(d) and 4(e).

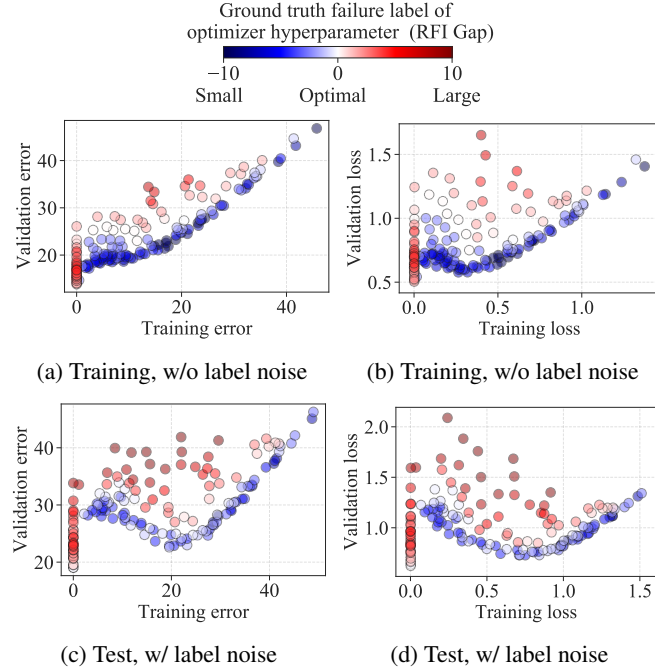


Figure 5. (Applying validation metrics such as training/validation error and loss to diagnose the failure source of models). (a)(b) Training set comprises models trained without label noise, (c)(d) test set comprises models trained with label noise.