

## Model Complexity:

There are many ways to measure the model complexity, here we discuss 3:

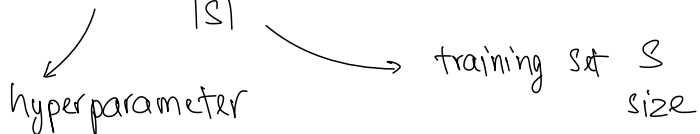
### Validation set

Using a "hold-out" set of data to measure the error rate of the learner on.

### Trade-off training error with "model complexity"

The potential function here is defined as  $\phi: \text{Trees} \rightarrow \mathbb{R}$

$$\phi(T) = \text{Training error on } S + \alpha \frac{\text{size}(T)}{|S|}$$



The goal is to minimize  $\phi$

### Minimal Description Length

This approach defines the best model is the one that compresses the data most efficiently.

Total Description Length = Bits to encode  $T$  + Bits to encode the errors

The goal is to minimize Total Description Length