Parameter Tuning
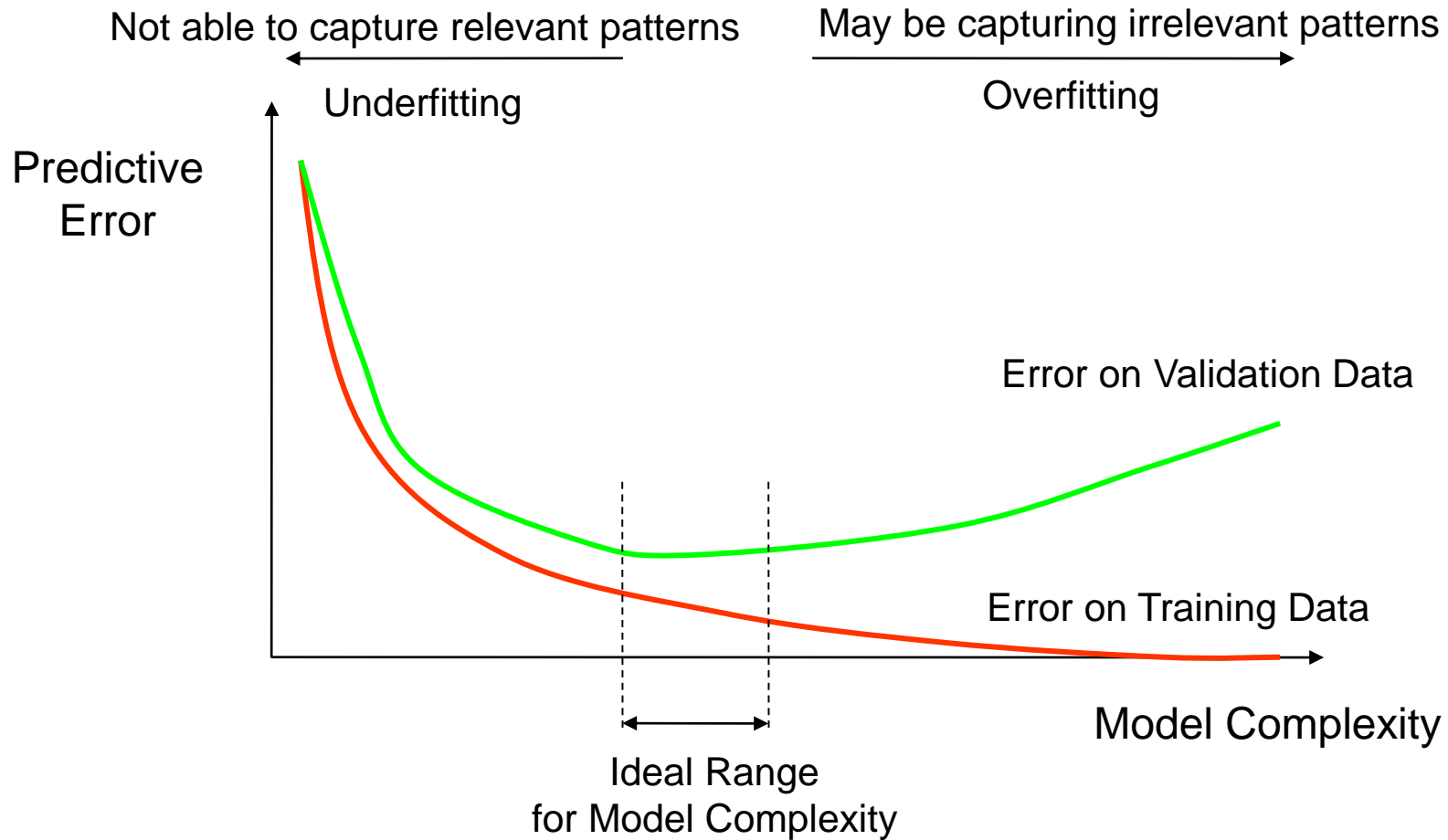
Building right model

- Overfitted model: If we build model that fits to the perfection on train data then it may not perform best on future unseen data.

- Diagnosis: Resubstitution(or train) error is much lower than cv error

- Why doesn't it perform best?
  - The train data may have noise observations
  - The train data may not have enough observations

Building right model

Parameter Tuning Problem

- Which parameter values do you select for building final model?

- Goal: Select the parameters of model that provides best performance on unseen(future) data.

Parameter Tuning Algorithm

- Define sets of model parameter values to evaluate

- For each parameter set do
  - Estimate the average error/accuracy of model using chosen resampling technique

- Choose the parameters that gives smallest error or highest accuracy

- Finally, build the model on entire data with the chosen parameters.