Model Evaluation

Ali Sharifi-Zarchi
Behrooz Azarkhalili
Peyman Naseri

CE Department
Sharif University of Technology

Fall 2022

VALIDATION

Main points why we evaluate the predictive performance of a model:

- 1. Want to estimate the generalization performance, the predictive performance of our model on future (unseen) data.

Main points why we evaluate the predictive performance of a model:

- 1. Want to estimate the generalization performance, the predictive performance of our model on future (unseen) data.
- 2. Want to increase the predictive performance by tweaking the learning algorithm and selecting the best performing model from a given hypothesis space.

Main points why we evaluate the predictive performance of a model:

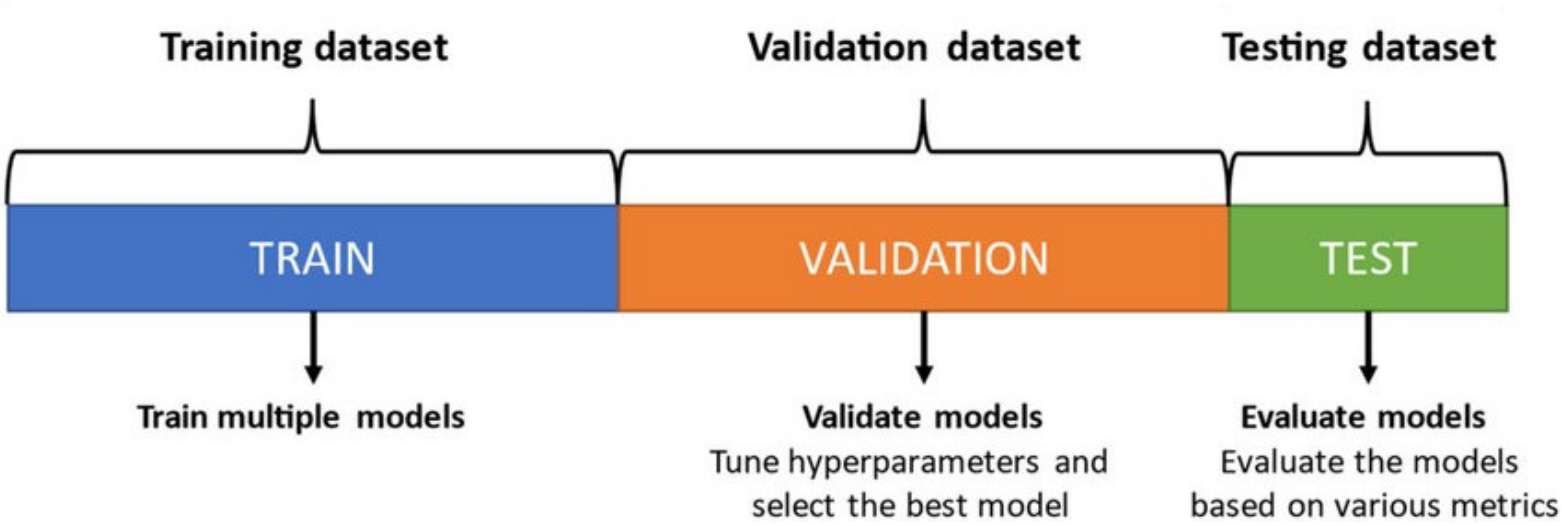
- 1. Want to estimate the generalization performance, the predictive performance of our model on future (unseen) data.
- 2. Want to increase the predictive performance by tweaking the learning algorithm and selecting the best performing model from a given hypothesis space.
- 3. Want to identify the ML algorithm that is best-suited for the problem at hand; thus, we want to compare different algorithms, selecting the best-performing one as well as the best performing model from the algorithm's hypothesis space

- Training set error is an optimistically biased estimator of the generalization error

- Training set error is an optimistically biased estimator of the generalization error
- Test set error is an unbiased estimator of the generalization error (test sample and hypothesis chosen independently)

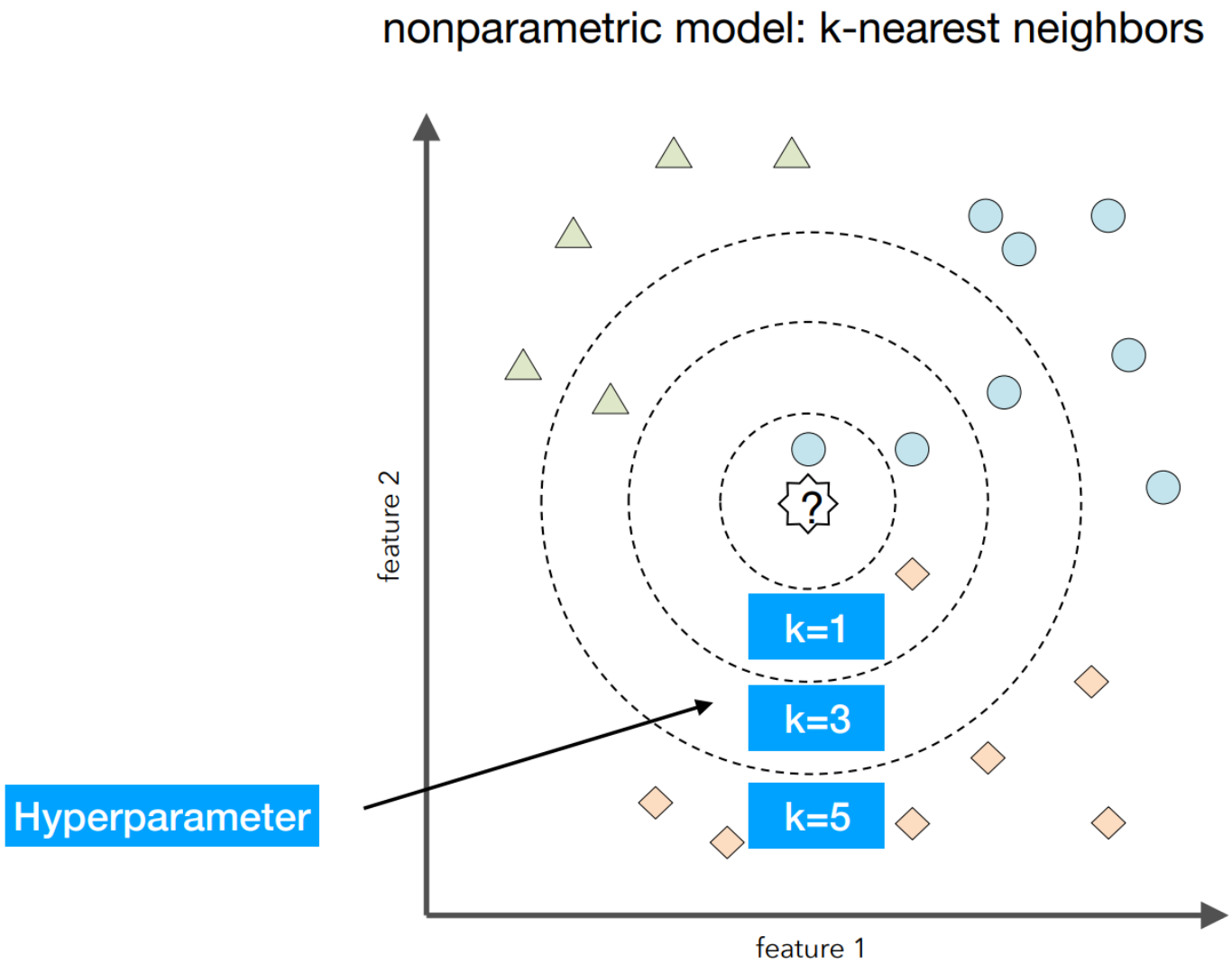
- Training set error is an optimistically biased estimator of the generalization error
- Test set error is an unbiased estimator of the generalization error (test sample and hypothesis chosen independently)
- (in practice, it is actually pessimistically biased; why?)

Validation

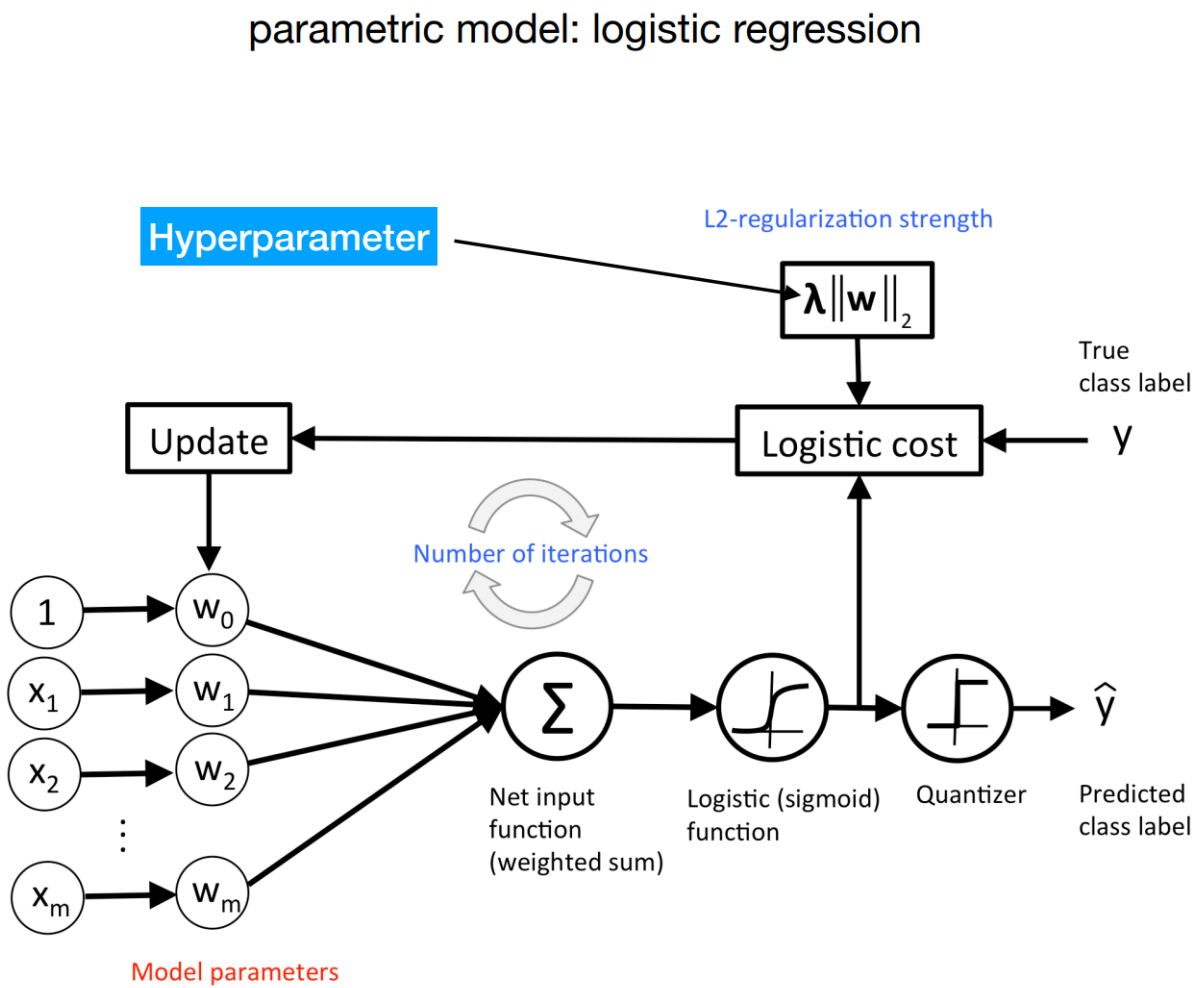

Training dataset
Validation dataset
Testing dataset
TRAIN
VALIDATION
TEST
Train multiple models
Validate models
Tune hyperparameters and select the best model
Evaluate models
Evaluate the models based on various metrics

Hyperparameters

Hyperparameters



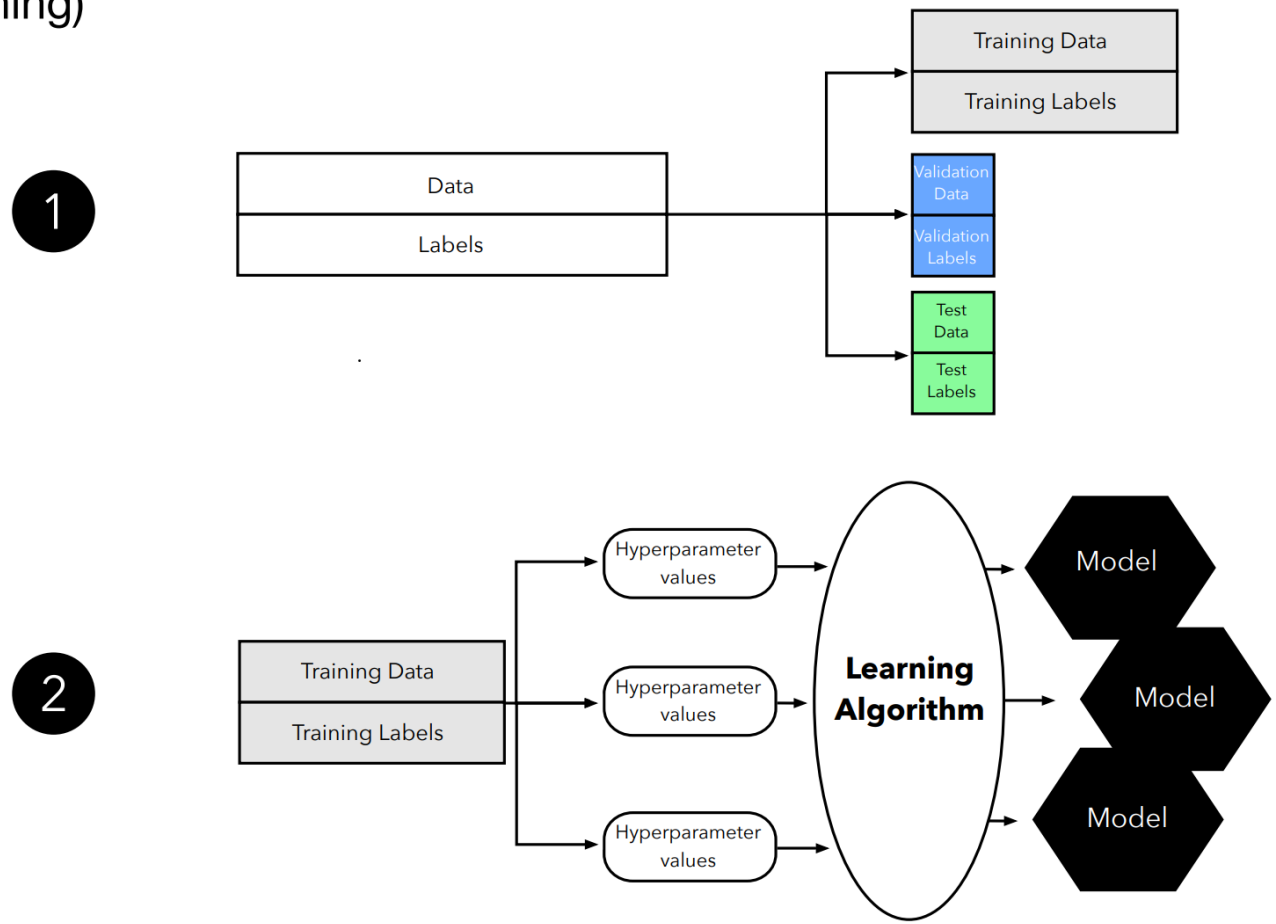
Hyperparameters



Holdout Validation

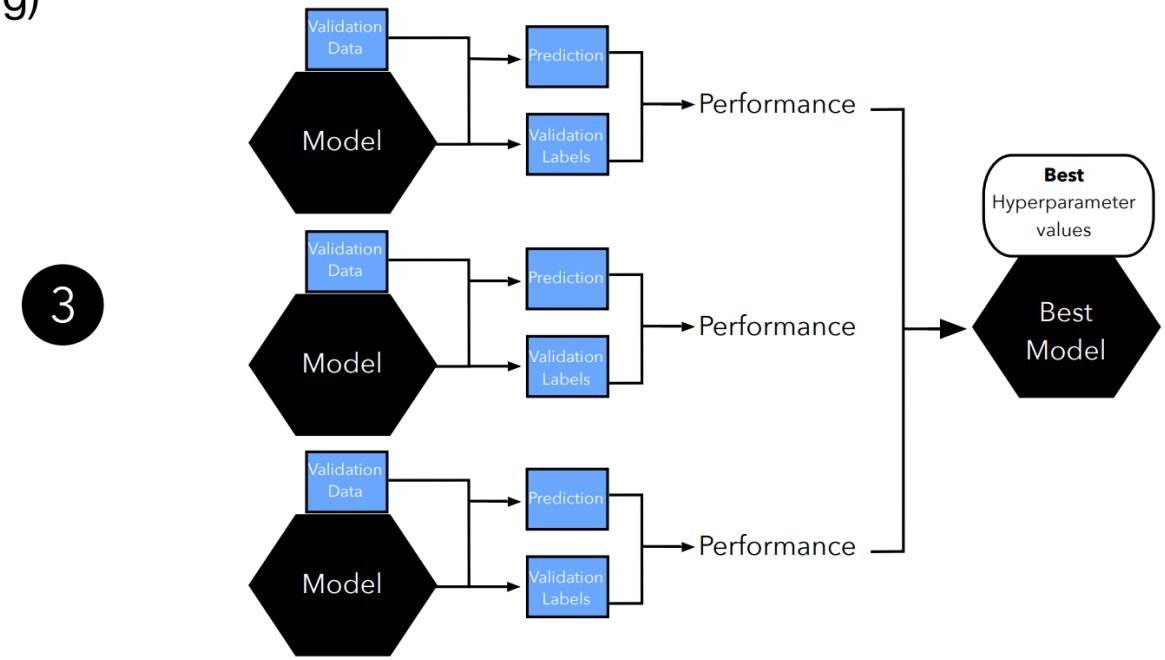
Holdout Validation

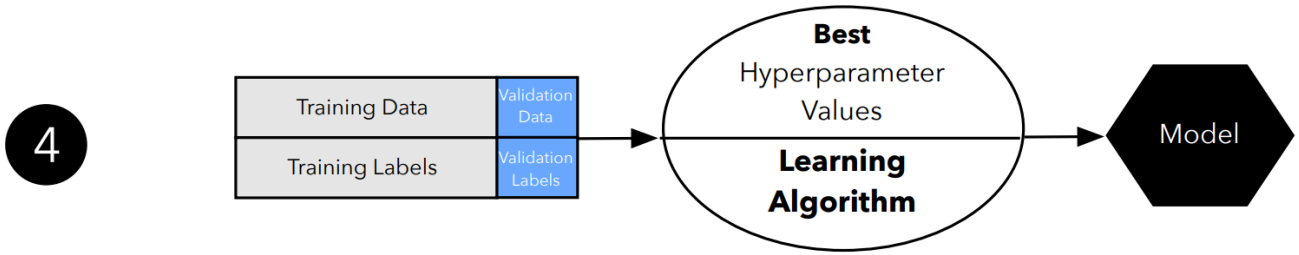
Holdout validation
(hyperparam. tuning)



Holdout Validation

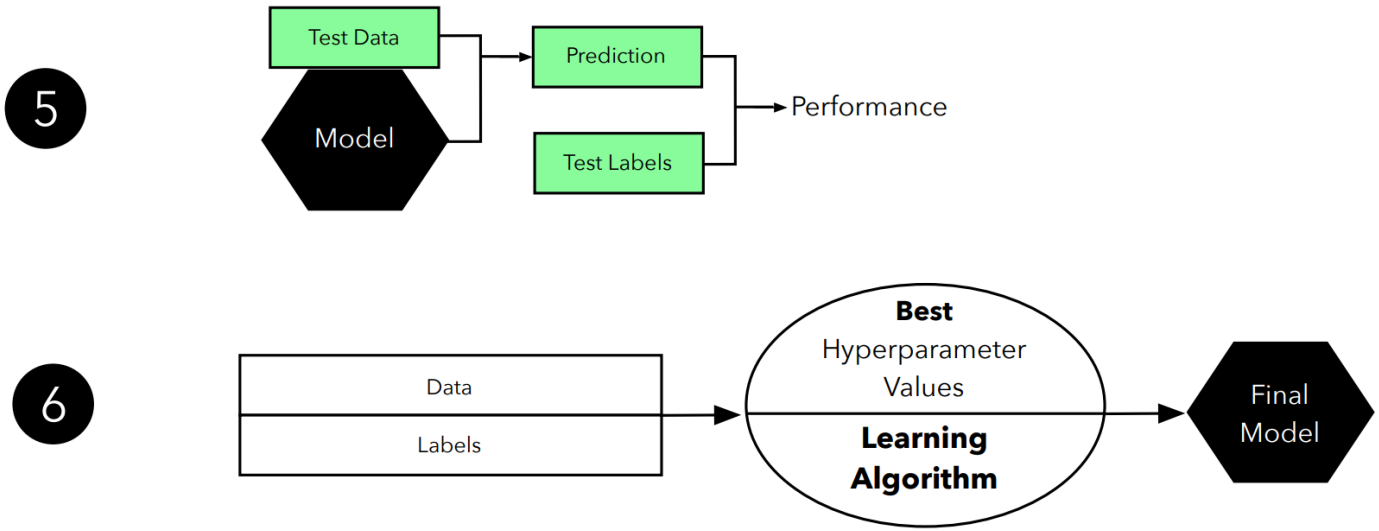
Holdout validation
(hyperparam. tuning)





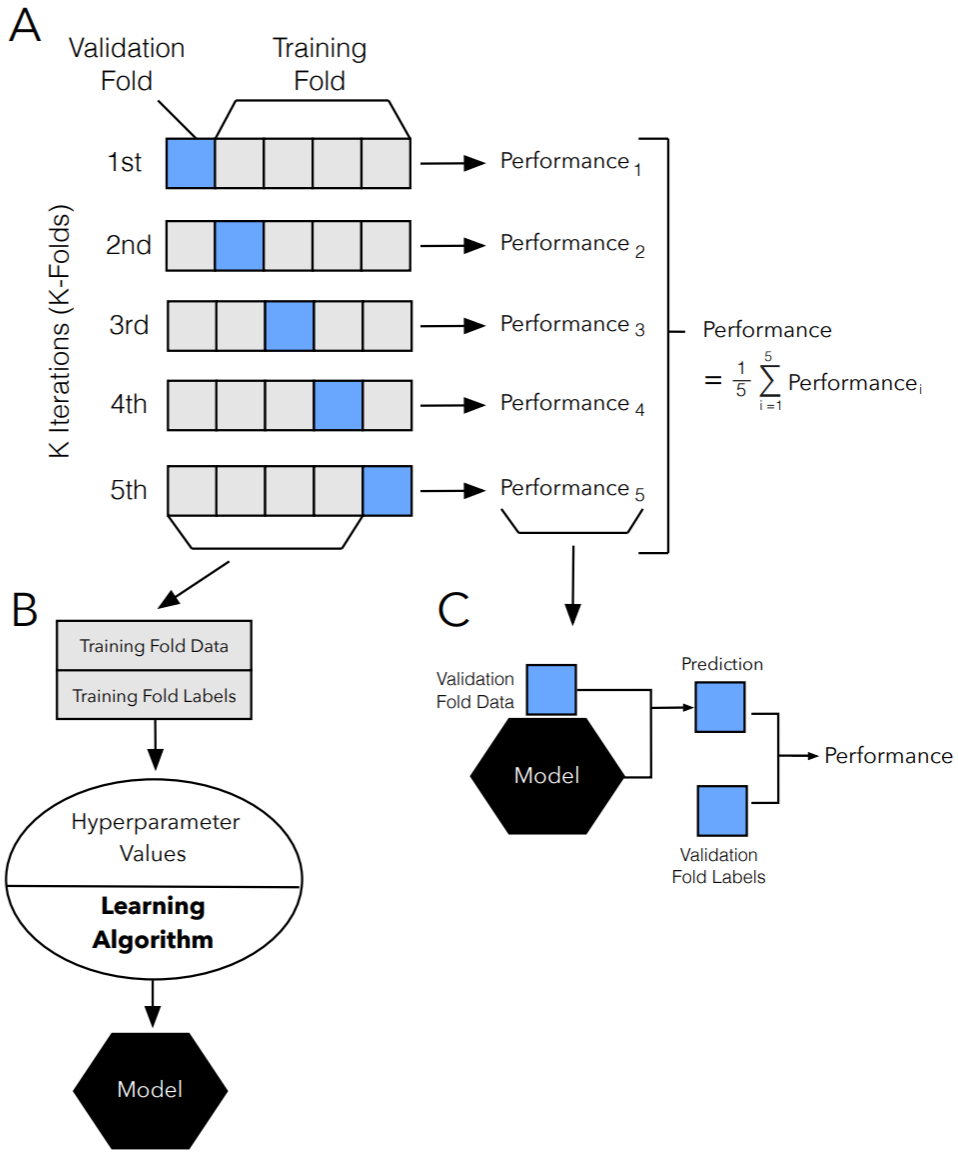
Holdout Validation

Holdout validation
(hyperparam. tuning)

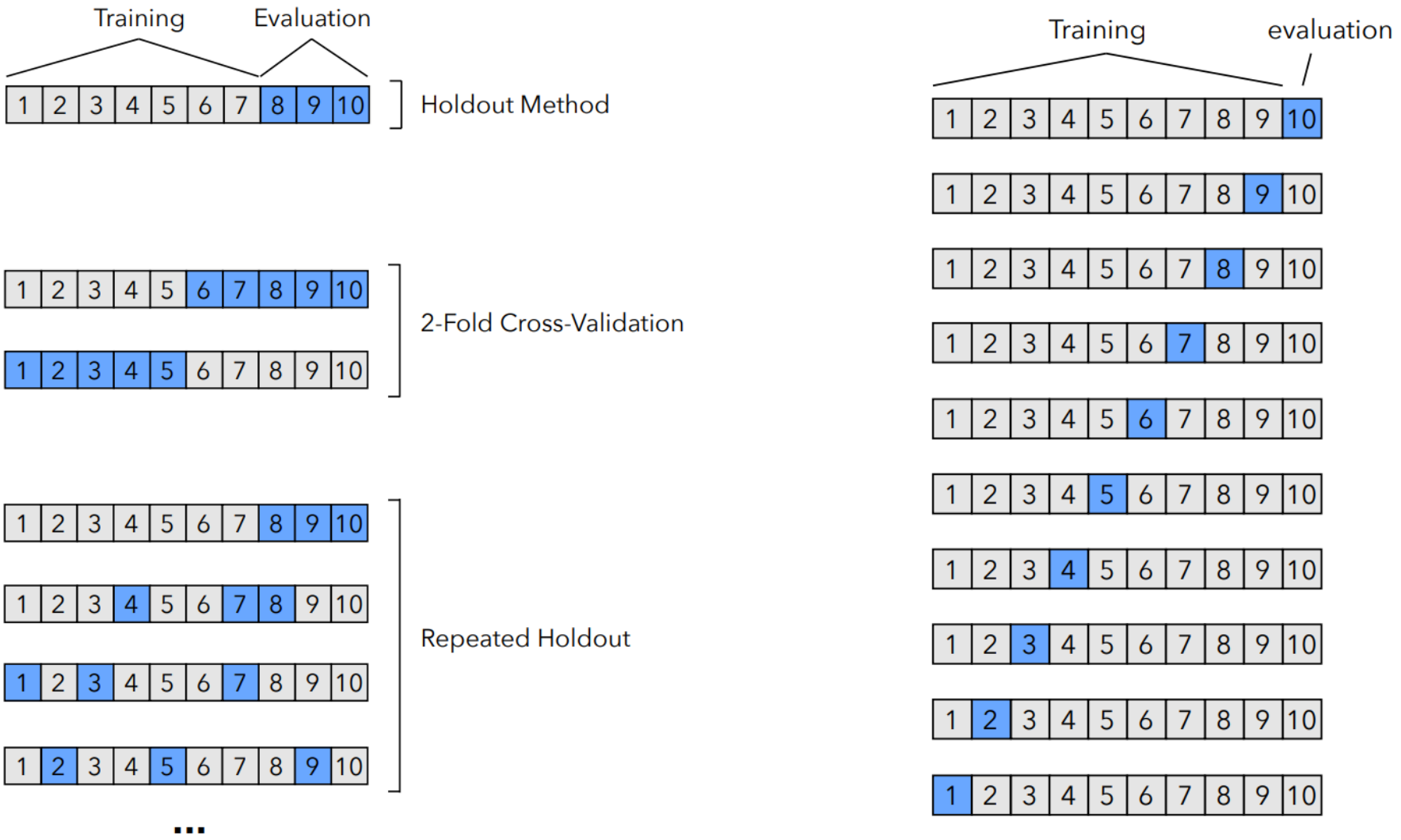


k-Fold Cross-Validation

k-Fold Cross-Validation



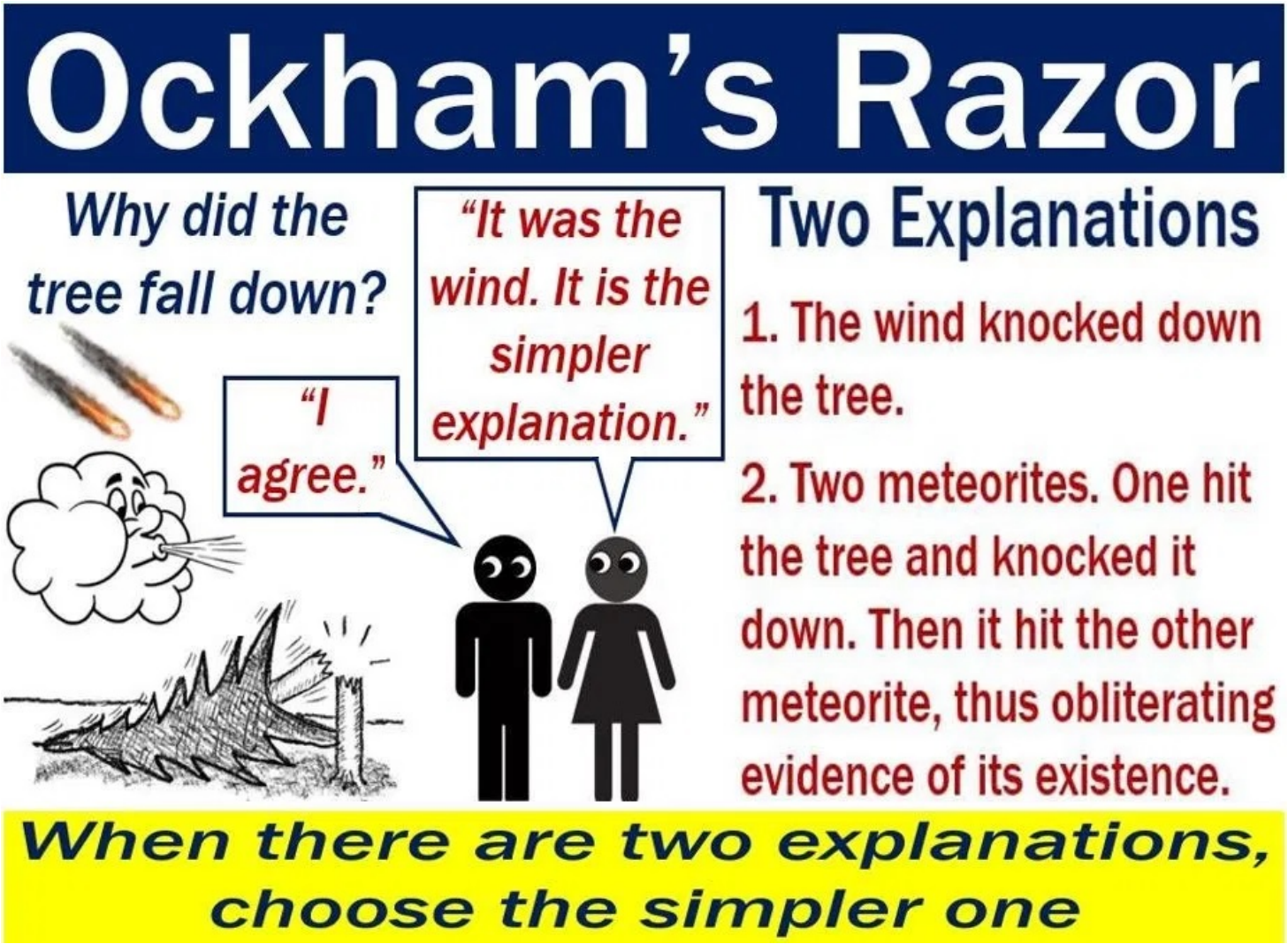
Types of validation



Model Selection

Model Selection

Model Selection



Thank You!

Any Question?