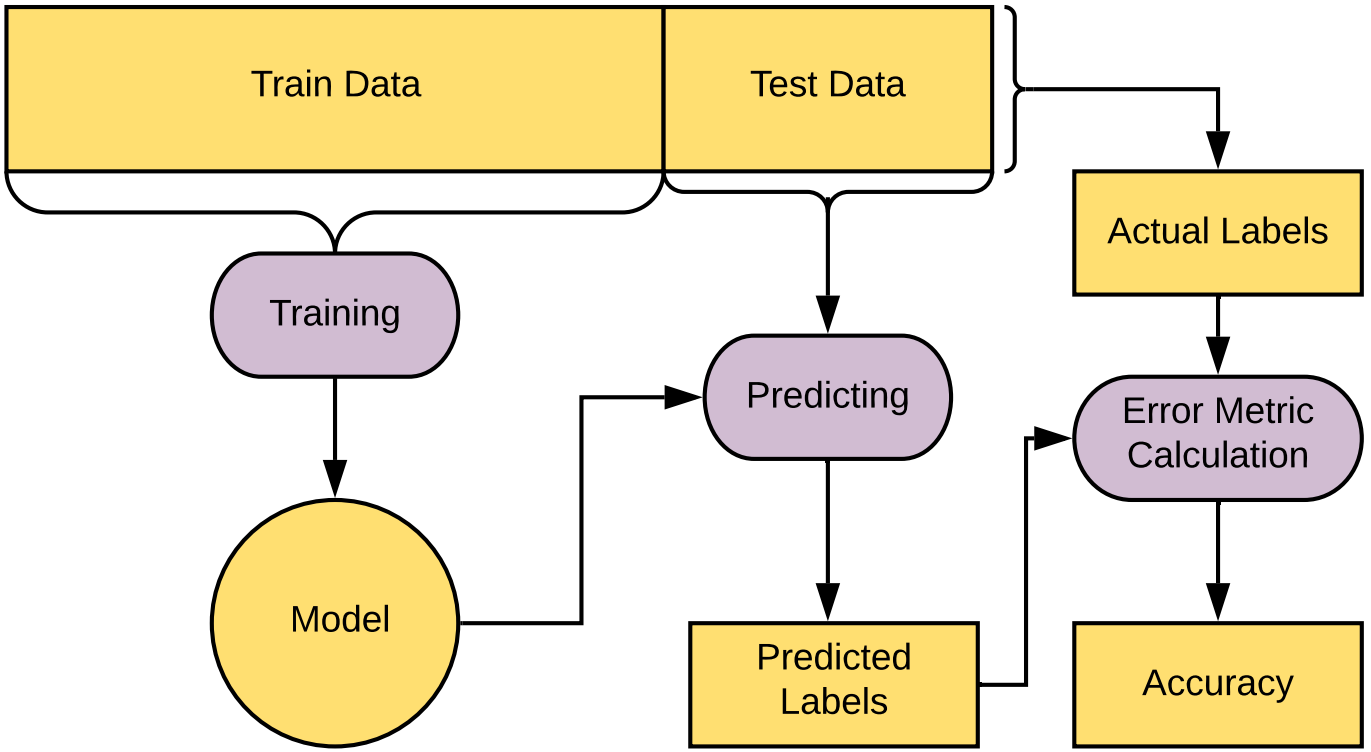
Cross-Validation

Nipun Batra and teaching staff

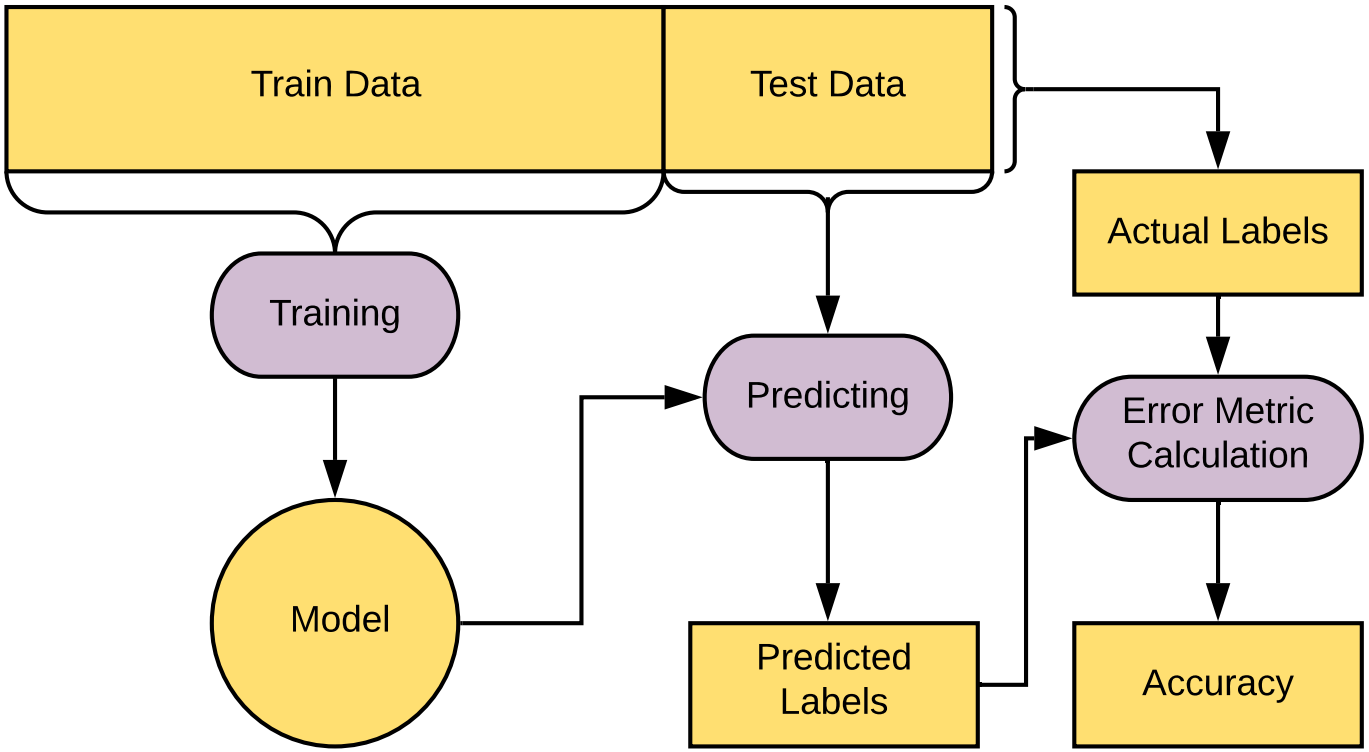
January 11, 2024

IIT Gandhinagar

Our General Training Flow

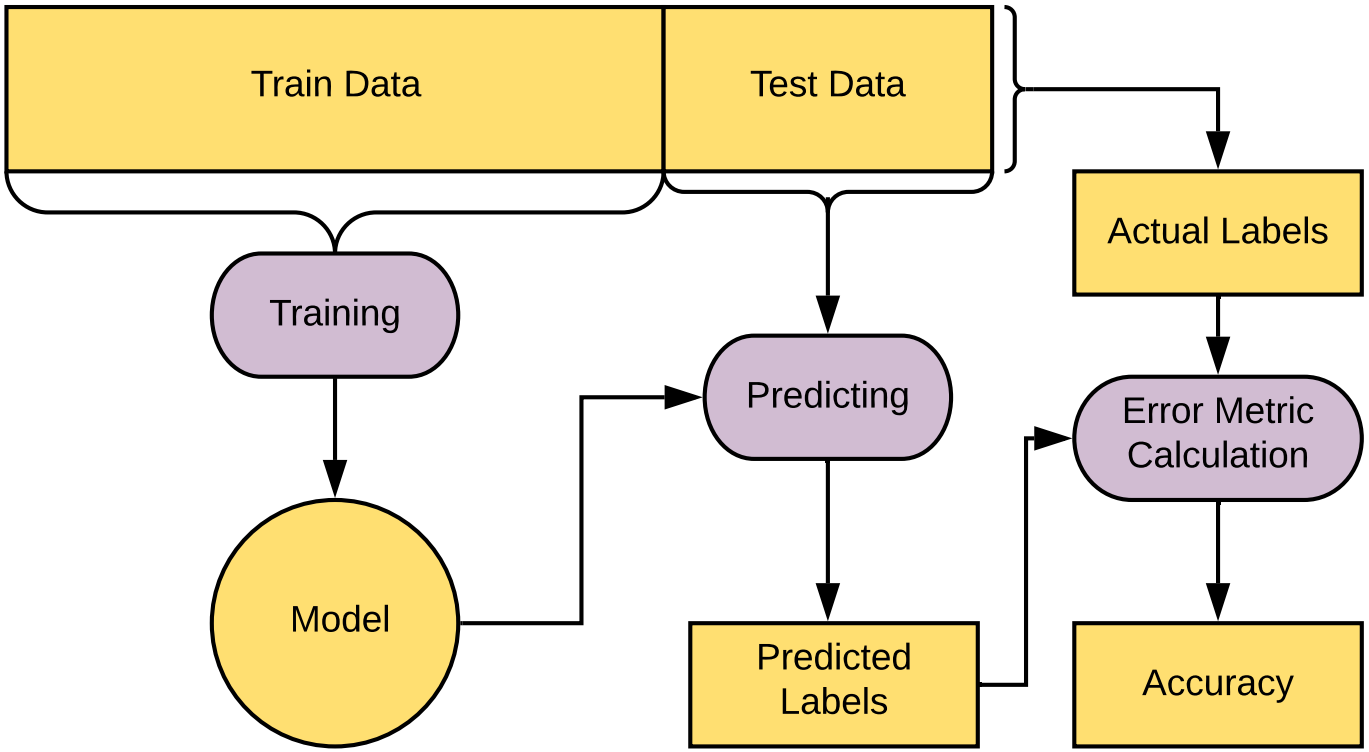


Our General Training Flow



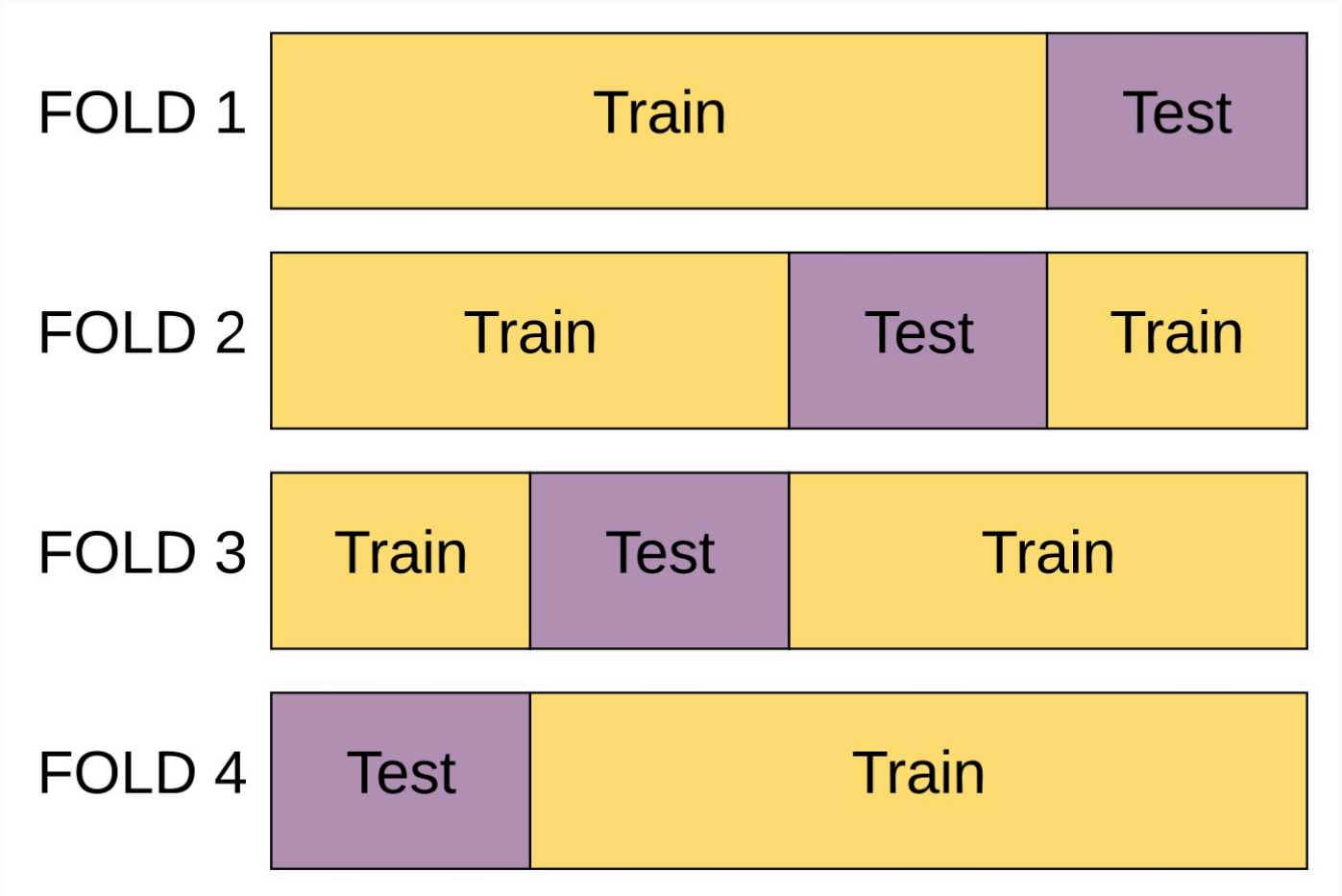
- Does not use the full dataset for training and does not test on the full dataset

Our General Training Flow

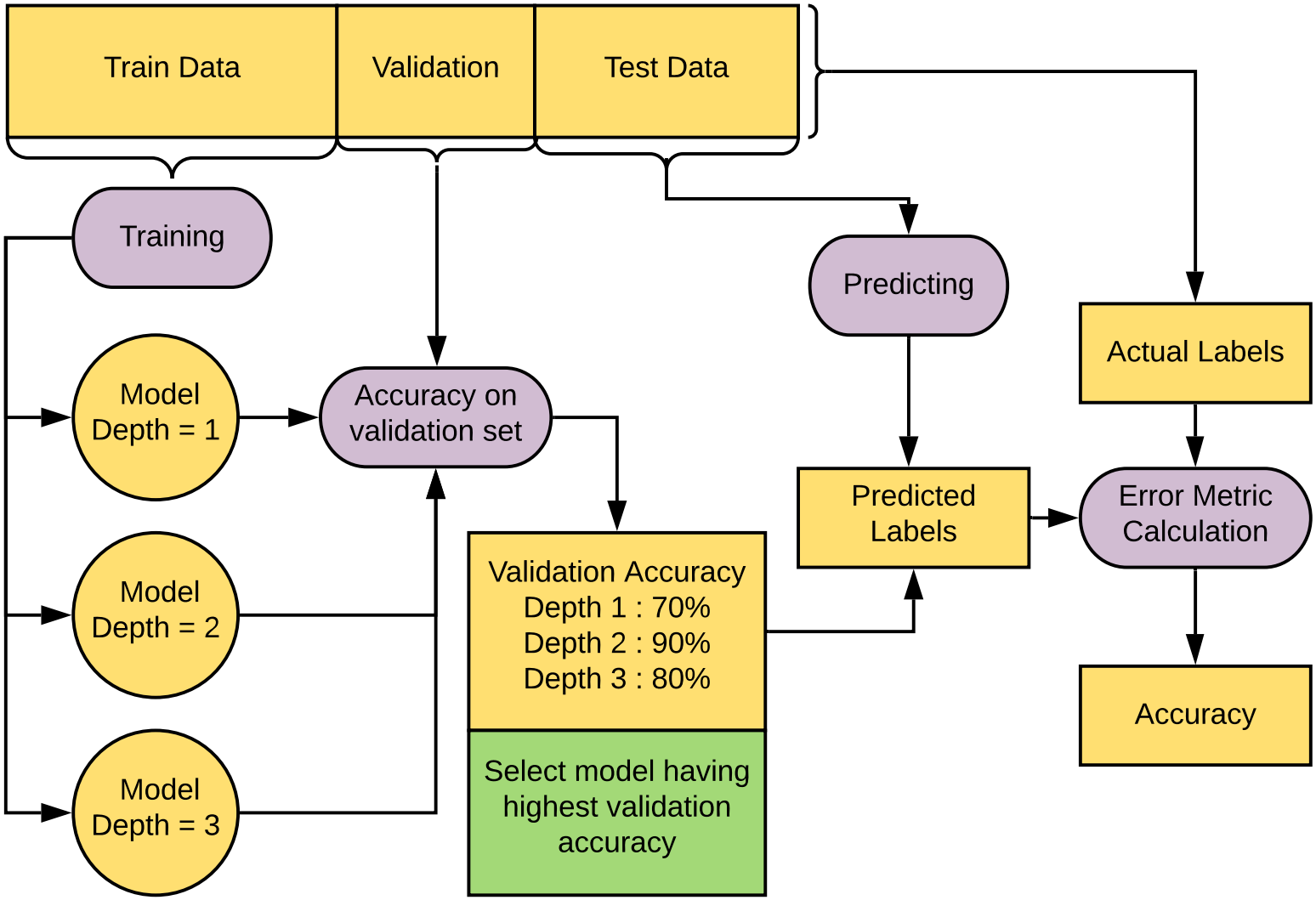


- Does not use the full dataset for training and does not test on the full dataset
- No way to optimise hyperparameters

K-Fold cross-validation: Utilise full dataset for testing



The Validation Set

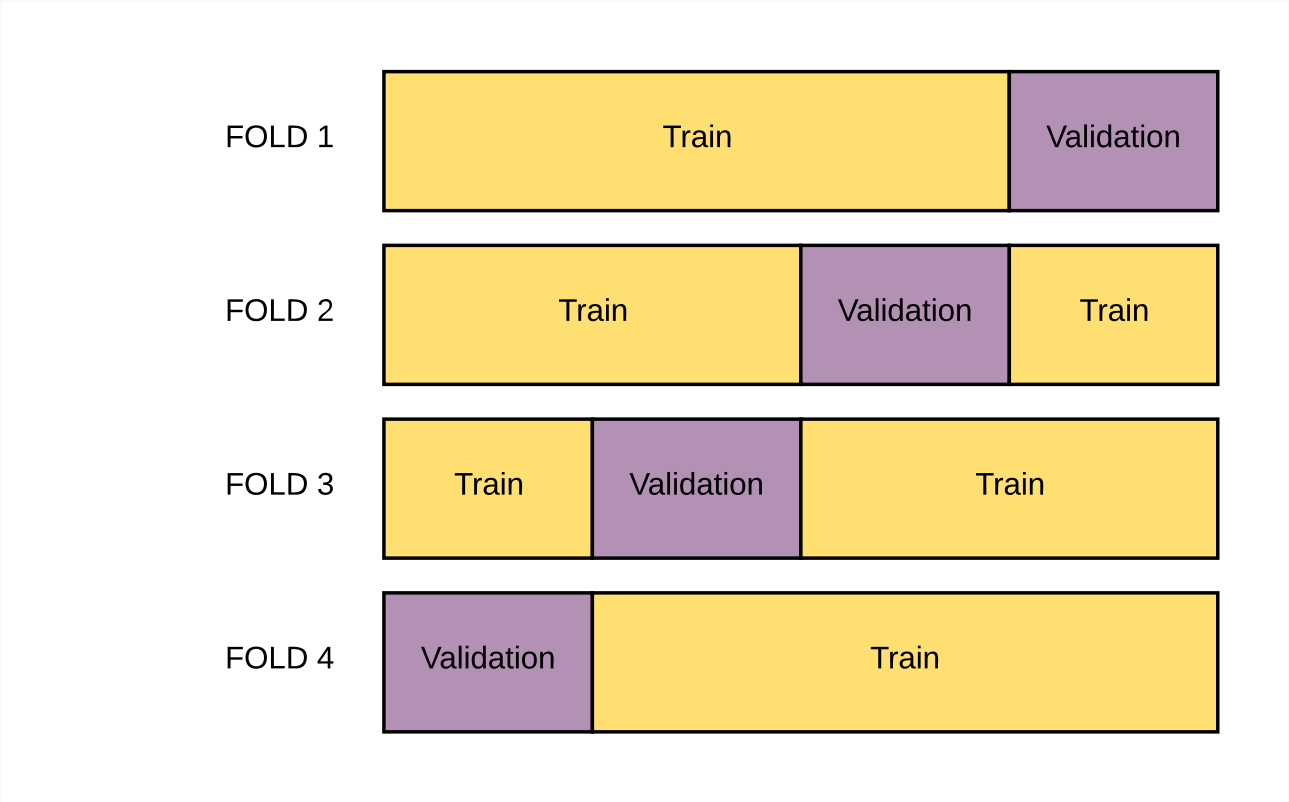


Nested Cross Validation

Divide your training set into K equal parts.

Cyclically use 1 part as "validation set" and the rest for training.

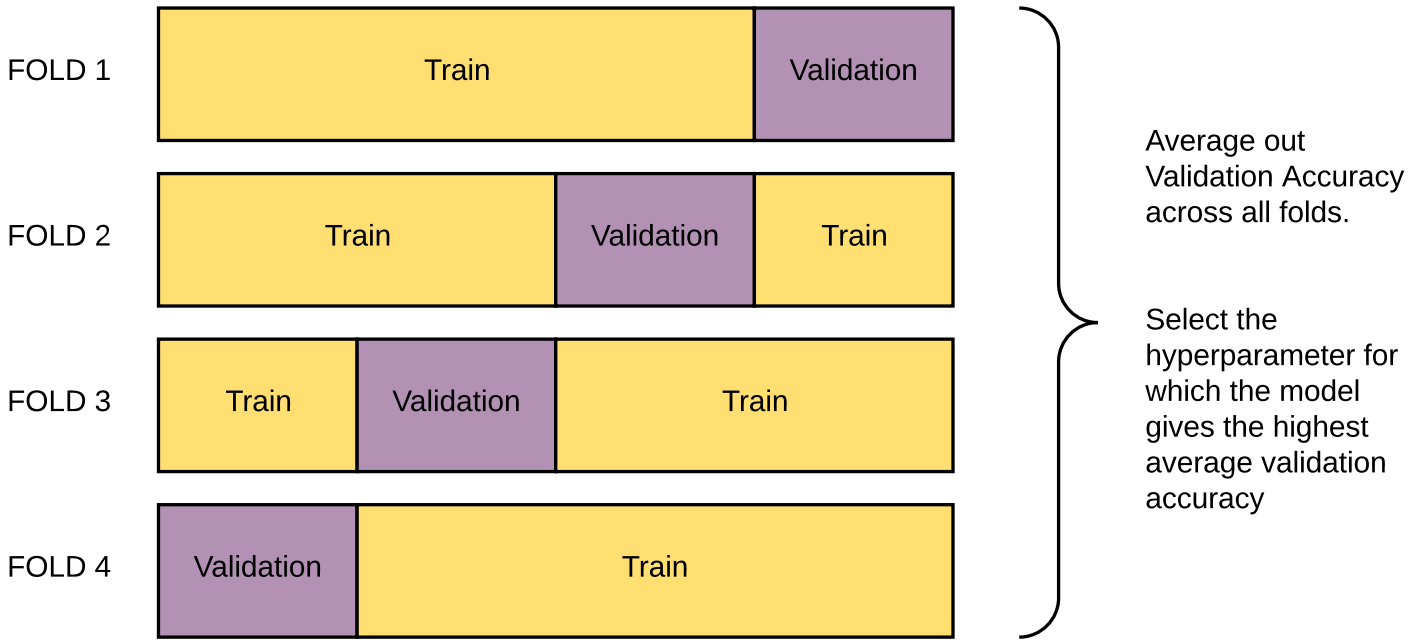
Here $K = 4$



Nested Cross Validation

Average out the validation accuracy across all the folds

Use the model with highest validation accuracy



Next time: Ensemble Learning

- How to combine various models?
- Why to combine multiple models?
- How can we reduce bias?
- How can we reduce variance?