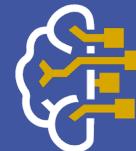


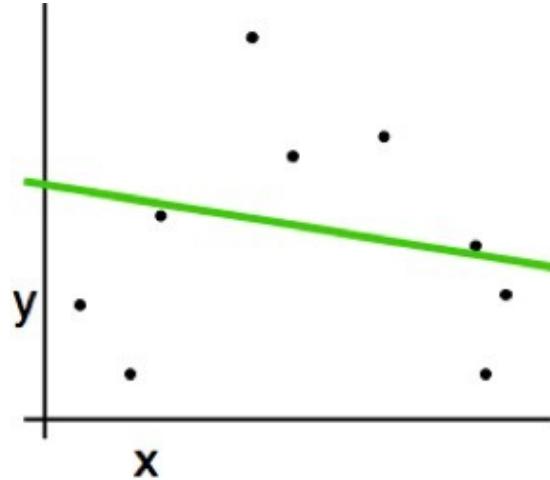
Hyperparameters

Lesson

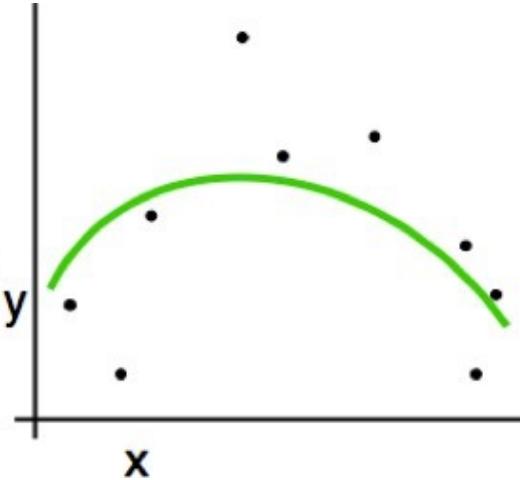


AI Academy

Which is the best?



(1)

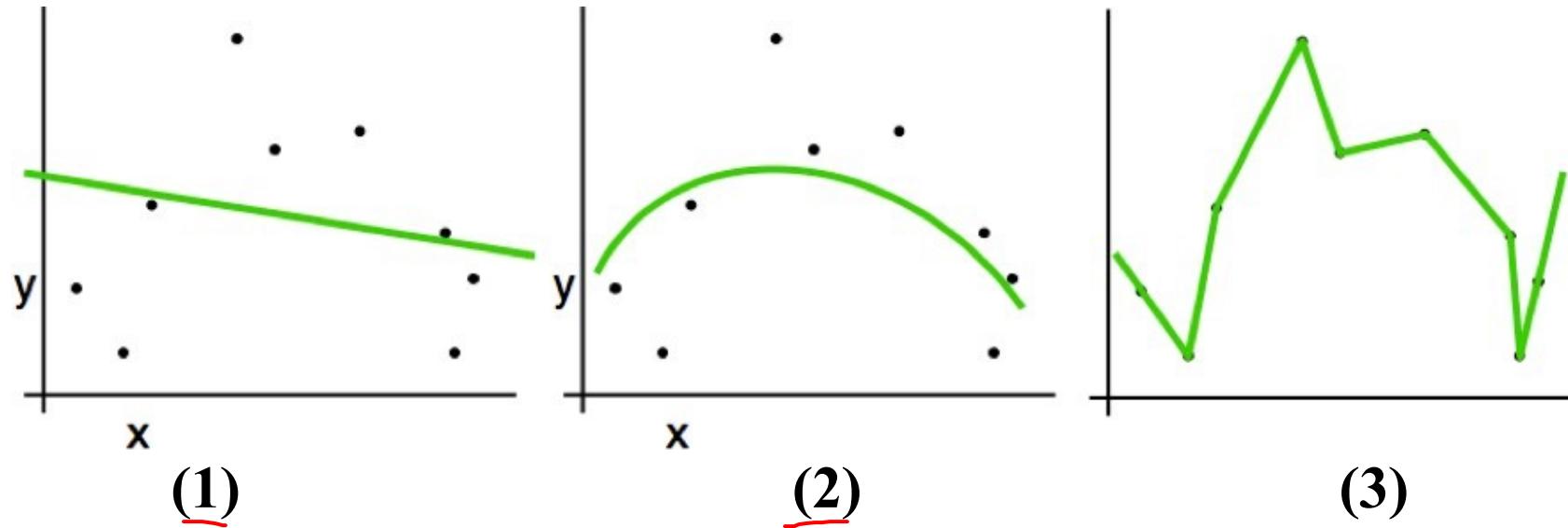


(2)



(3)

Hyperparameters

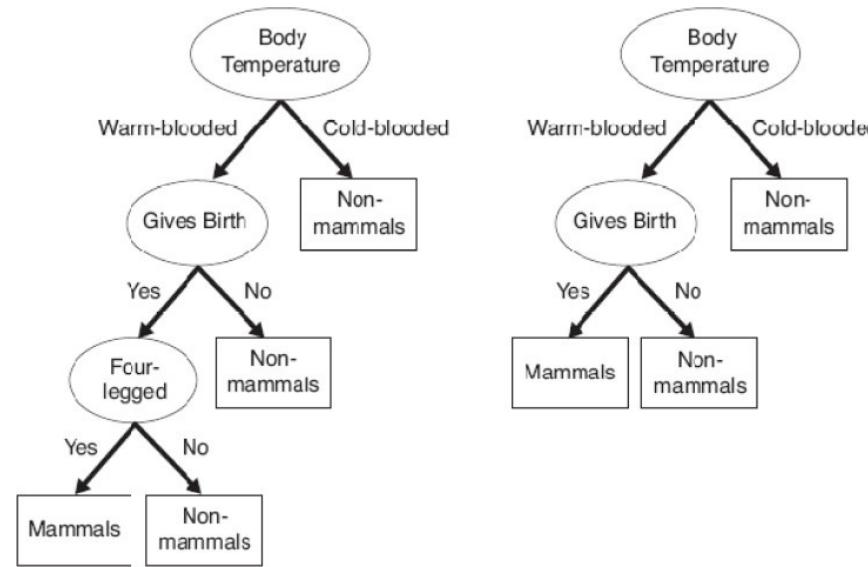


Hyperparameter: Parameter of our model that is set before training.

Example: The degree of the polynomial in a regression model.

This is different than a model parameter, which is fit during training.

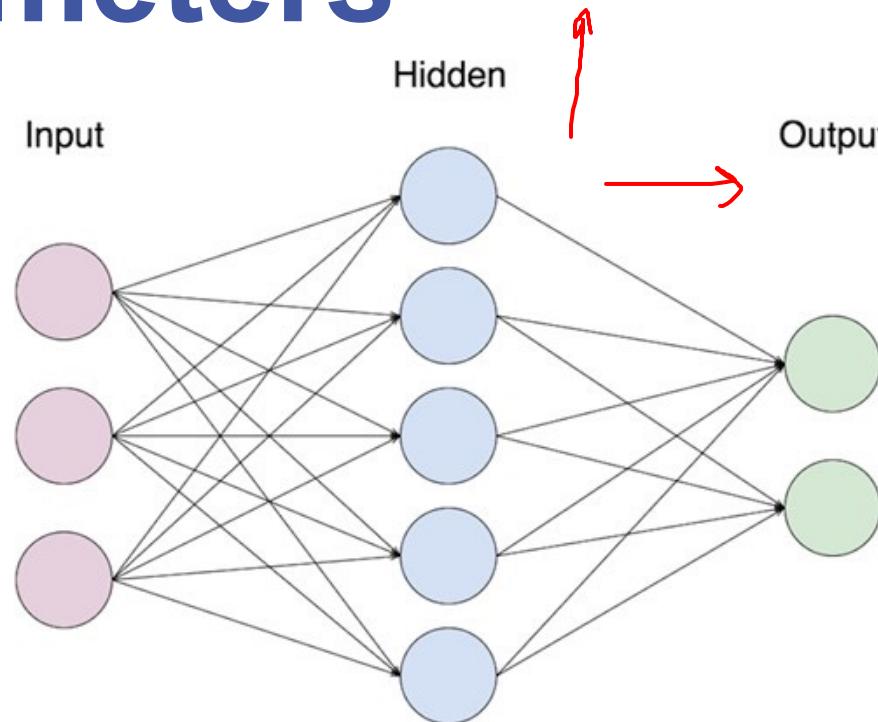
Hyperparameters



Hyperparameter: Parameter of our model that is set *before* training.

Example: The maximum depth of a decision tree (pre-pruning)

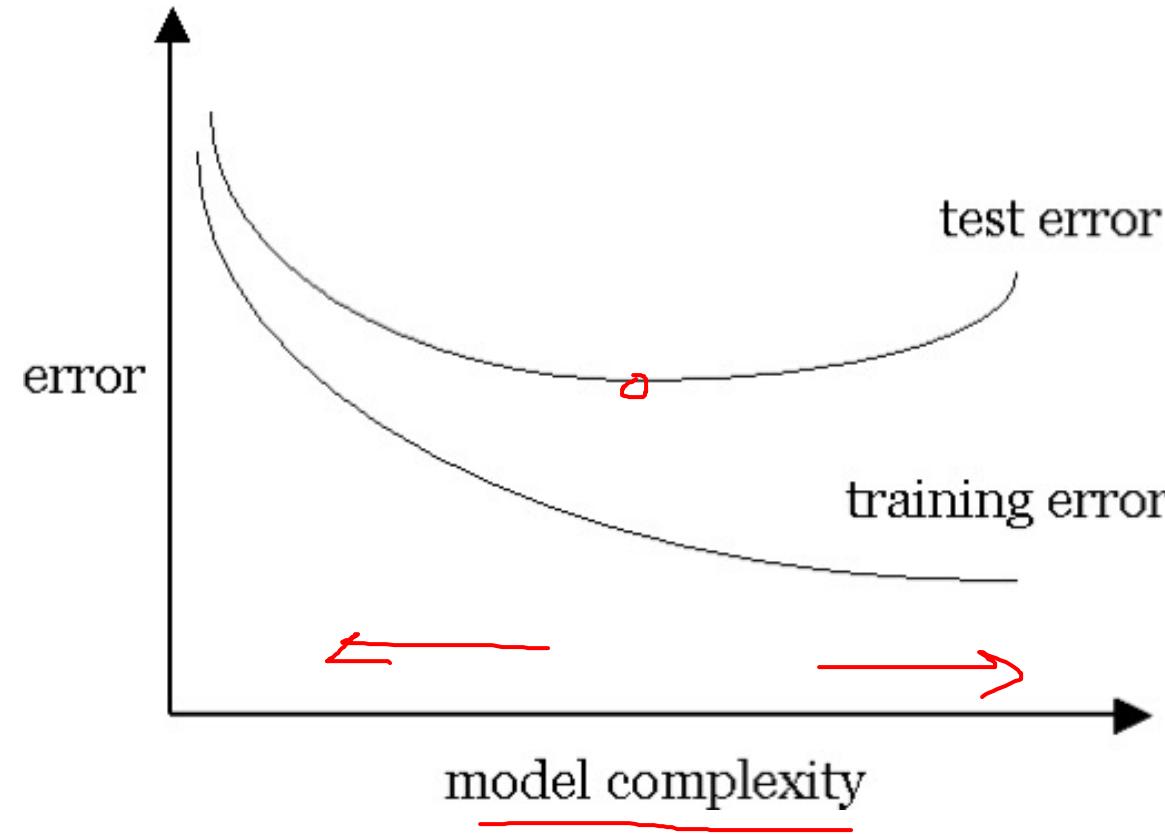
Hyperparameters



Hyperparameter: Parameter of our model that is set *before* training.

Example: The number of nodes or hidden layers in a neural network.

Hyperparameters & Overfitting



We call choosing a suboptimal model:
Underfitting or Overfitting.

A learning system: basic cycle

1) Data: $D = \{d_1, d_2, \dots, d_n\}$

2) Choose a learner:

Select a model with parameters
(e.g. Decision Tree).

Parameter: A model value learned in training.

3) Choose the objective function:

E.g. Training Error.

4) Learning:

Train the model to optimize the objective function.

A learning system: basic cycle

1) Data: $D = \{d_1, d_2, \dots, d_n\}$

2) Choose a learner:

Select a model with parameters
(e.g. Decision Tree).

Parameter: A model value learned in training.

Repeat step 4 for each set of hyperparameters.

3) Choose the objective function:

E.g. Training Error.

4) Learning:

Train the model to optimize the objective function.

5) Select hyperparameters

that optimize the objective function on the validation set.



Comparing Hyperparameter Values

What dataset do we use to compare hyperparameters?

- **Training Dataset**: Would lead to overfitting. 
- **Test Dataset**: Would be “cheating”! We cannot “peak” at our test data to find the best hyperparameters.
 - Poor estimate of generalization error.

Training, Validation, and Test Sets

Divide the data set into three distinct parts:

- ✓ **Training Dataset:** Used to train the model parameters (fit parameters).
- ✓ **Test Dataset:** Used for final evaluation of the tuned model, to estimate generalization error.
- **Validation Dataset:** Used to select hyperparameters, estimates model performance on the test dataset.

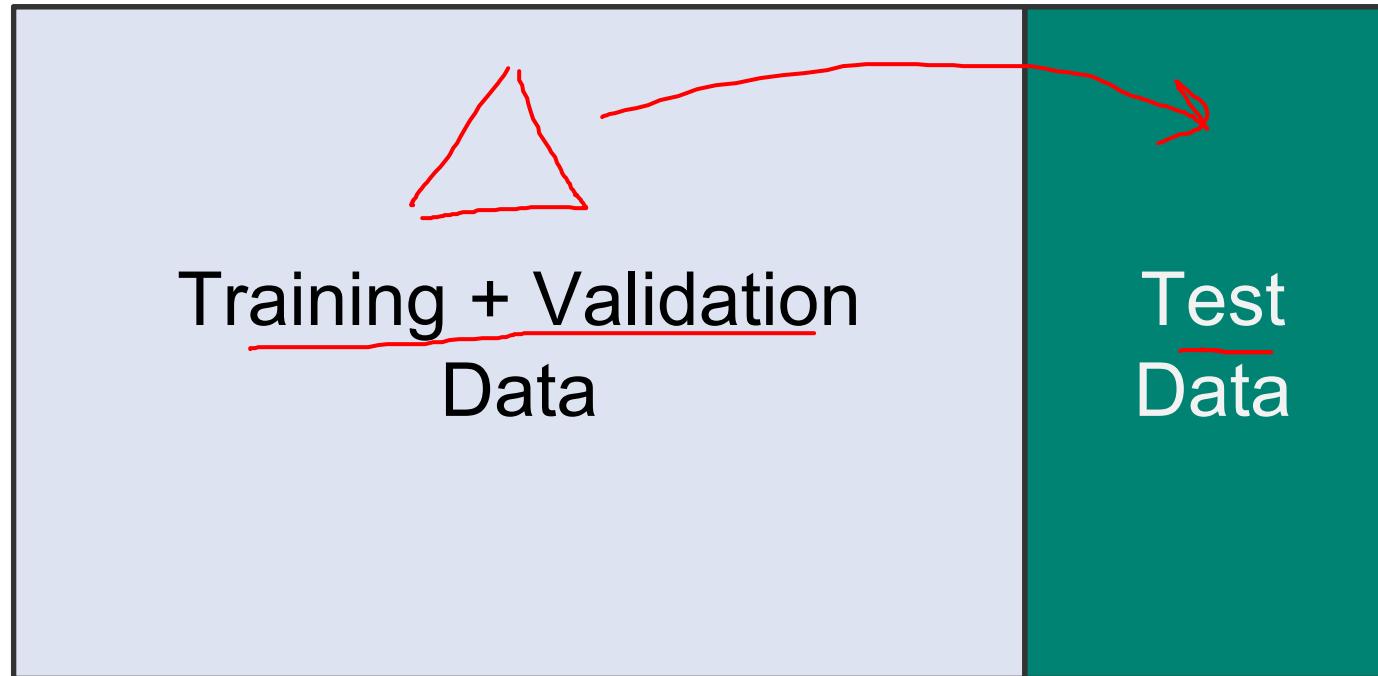
Selecting a Validation Dataset: Holdout

- Split data into train/validation/testing (e.g. 50%, 25%, 25%).



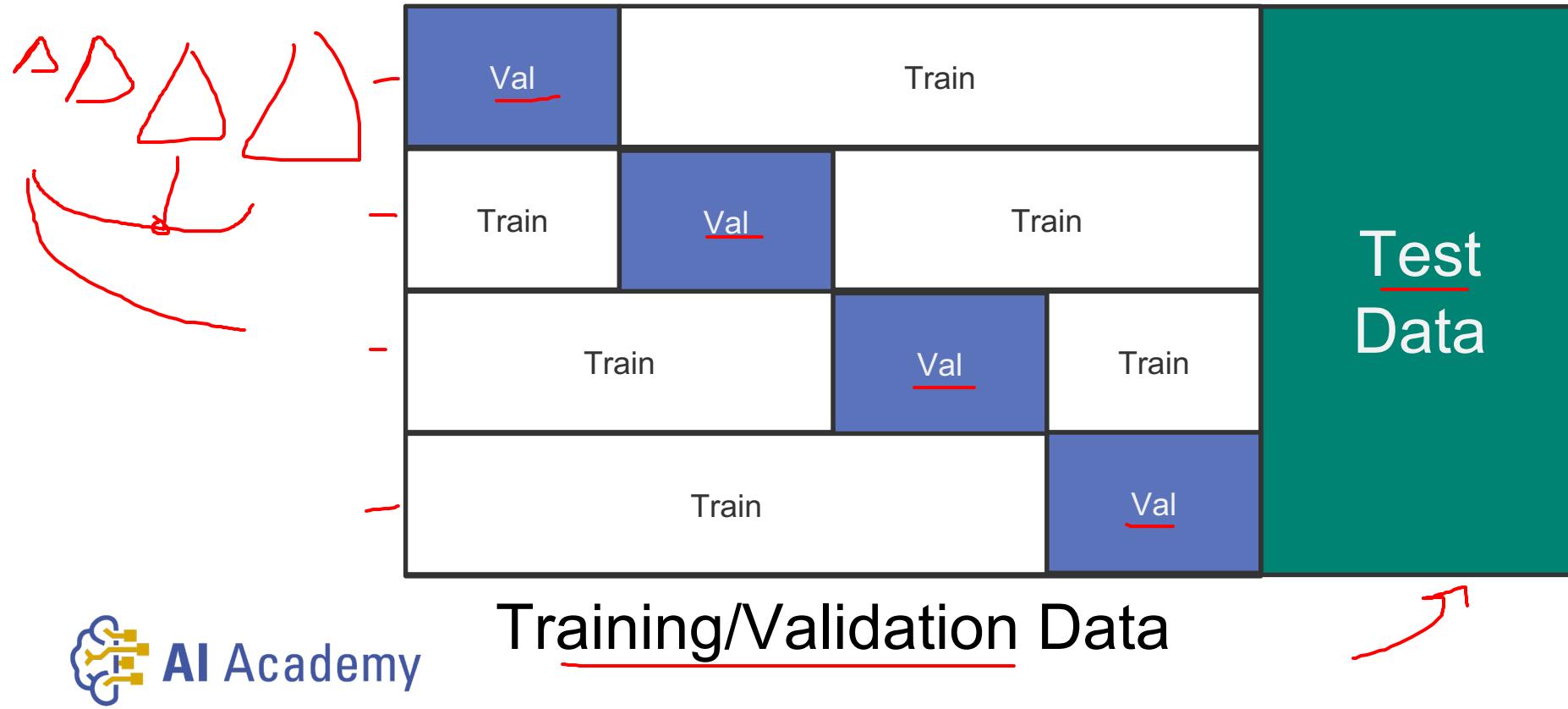
Selecting a Validation Dataset: Holdout

- Optional: After selecting hyperparameters, combine training + validation and use best value.



Selecting a Validation Dataset: CV

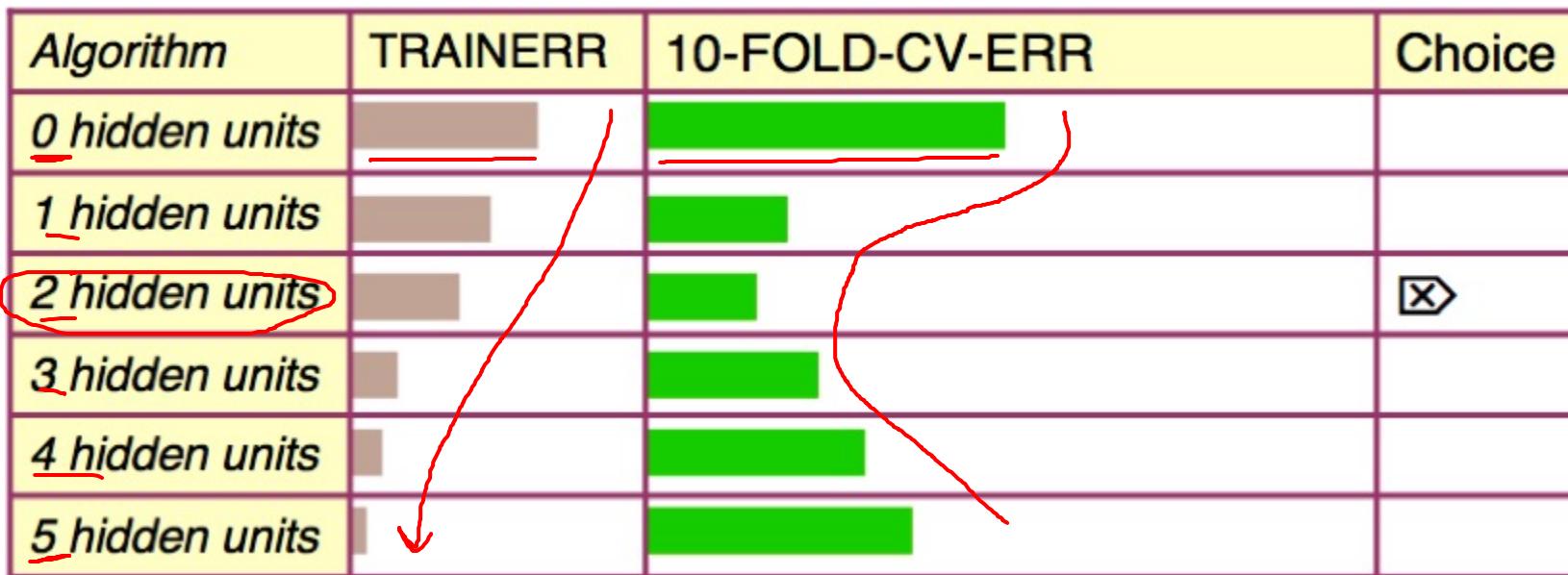
- Split into Training and Test, then do CV *within* training data.



CV-based Model Selection

Example: Choosing number of hidden units in a one-hidden-layer neural net.

- Step 1: Compute 10-fold CV error for six different models:

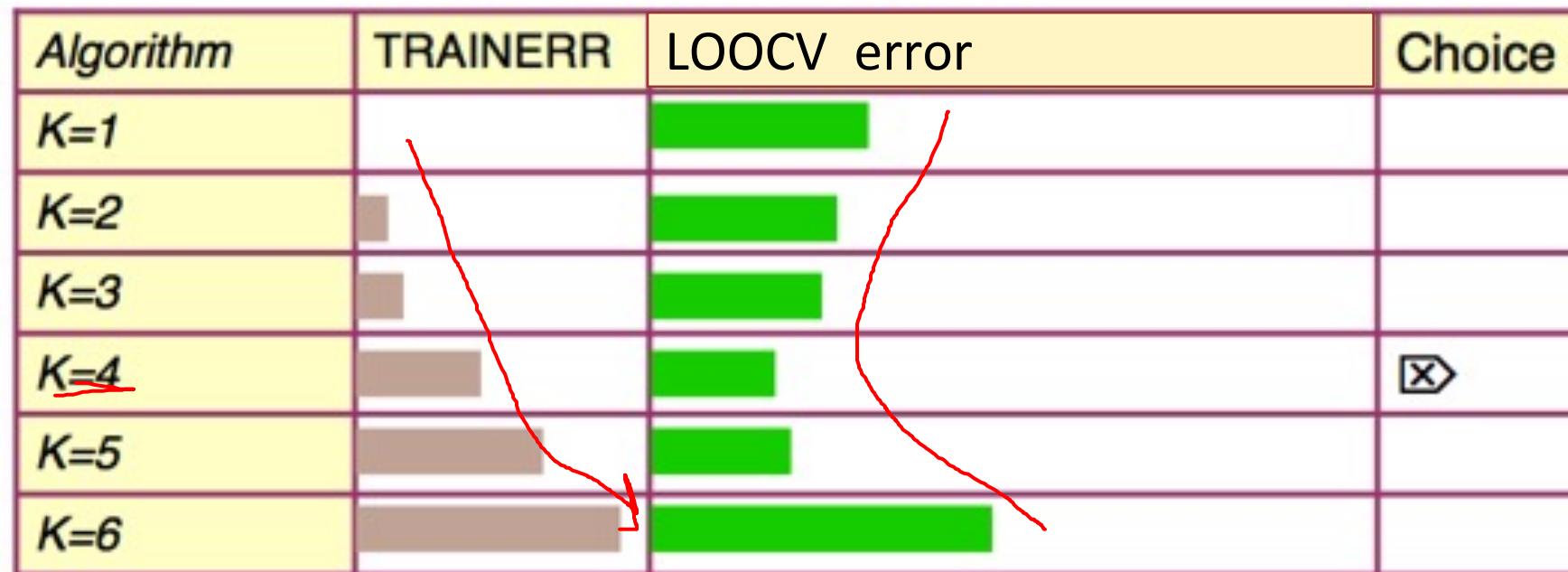


- Step 2: Whichever model class gave best CV score: train it with all the data, and that's the final predictive model.

CV-based Model Selection

Example: Choosing “k” for a k-nearest-neighbor (KNN) classifier.

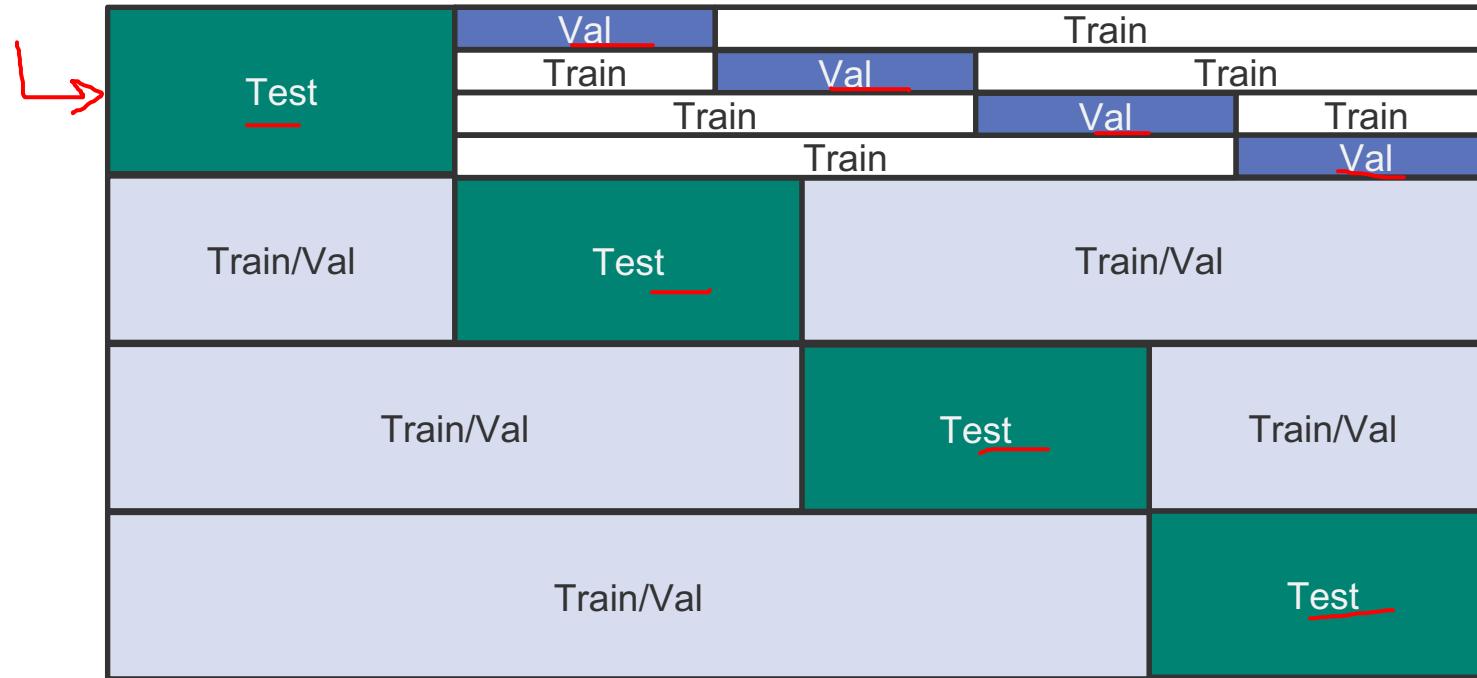
- Step 1: Compute LOOCV error for six different models:



- Step 2: Whichever model class gave best CV score: train it with all the data, and that's the final predictive model.

Selecting a Validation Dataset: Nested CV

- Do CV for training/test split *and* training/validation splits.



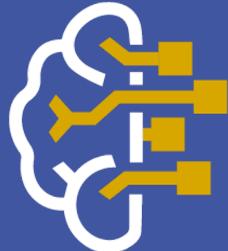
Learning Objectives: Hyperparameters

You now should be able to:

- Explain what a hyperparameter is, give examples, and understand how hyperparameters are learned/tuned.



AI Academy
NC STATE



AI Academy

go.ncsu.edu/aiacademy

NC STATE UNIVERSITY

Exercise

A researcher trains a k-NN model using a training dataset with 100 instances. She uses 4-fold CV *within* that training dataset to tune the hyperparameter k , choosing between 1, 3, and 5. She then evaluates the tuned model on a separate test dataset.

How many *total* k-NN models will be built, including those used for hyperparameter tuning?

- A) 100
- B) 4
- C) 12
- D) 13