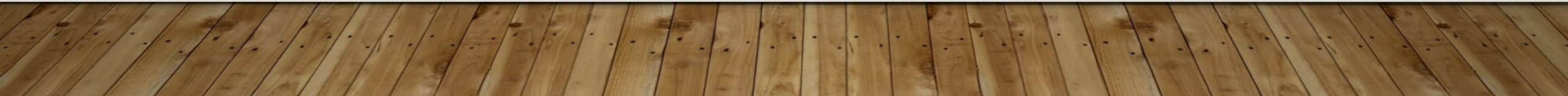


# GRIDS AND PIPES

---



# HYPERPARAMETERS

---

- We looked at using hyperparameters to make accurate models.
- Each algorithm has different hyperparameters.
- Hyperparameters are variables set ahead of time that control the training process.
- We can't evaluate optimal HP ahead of time, we need to set them and see the results.
  - Trial and error!

# GRID SEARCH

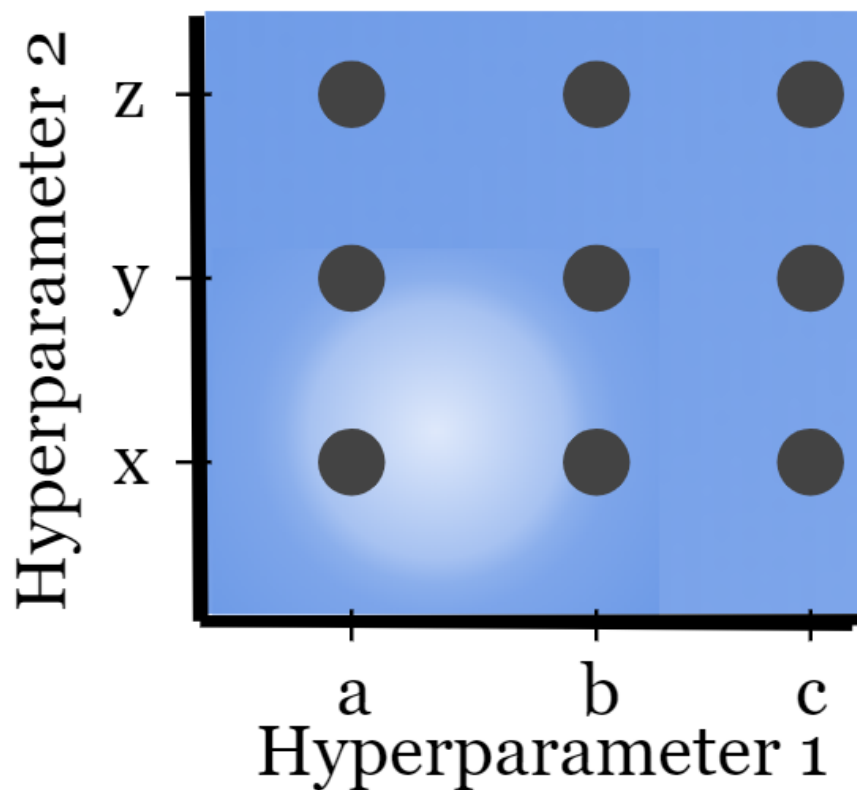
---

- We could loop through different combinations of HPs and collect scores.
  - Complex and clumsy – multiple nested loops.
- Sklearn provides a simplification – GridSearchCV.
- Grid search takes a list of HP values we want to try, checks every combination, and returns the best model.
- Same result of looping through each value.
- Also a randomized version – for when we have no idea what HP values may work.
- CV = cross validation....

# Grid Search

Pseudocode

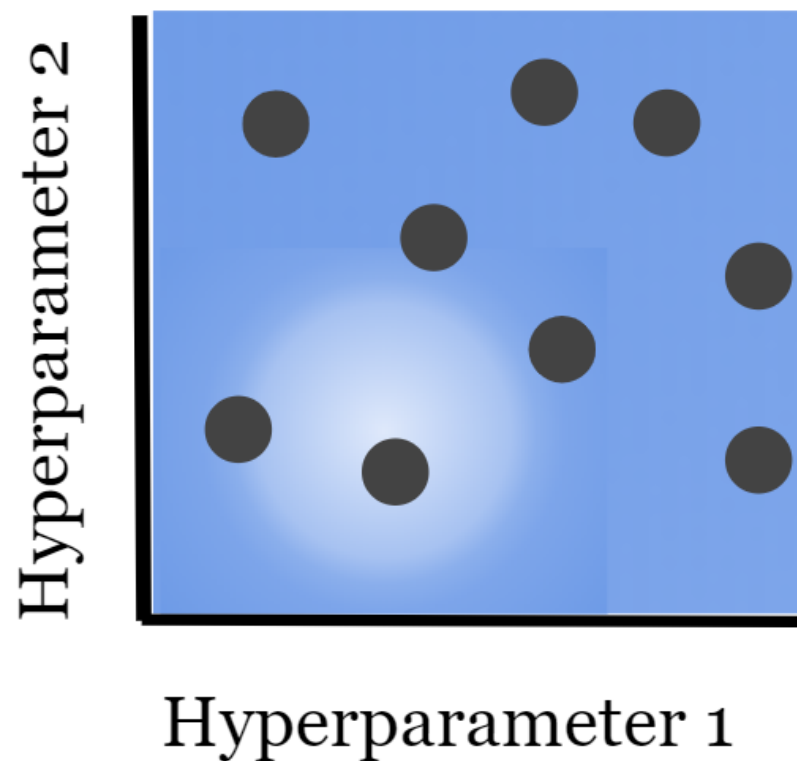
```
Hyperparameter_One = [a, b, c]  
Hyperparameter_Two = [x, y, z]
```



# Random Search

Pseudocode

```
Hyperparameter_One = random.num(range)  
Hyperparameter_Two = random.num(range)
```



# TRAIN-TEST SPLIT

---

- We are used to splitting data when building models:
  - Training data (~70%) used to create the predictive model.
  - Testing data (~30%) used to evaluate the accuracy of the model.
- This split allows us to accurately assess the accuracy on “new” data.
- Problem – random splits in the data can lead to variance in results:
  - Think of drastically different trees we get when data is split differently.

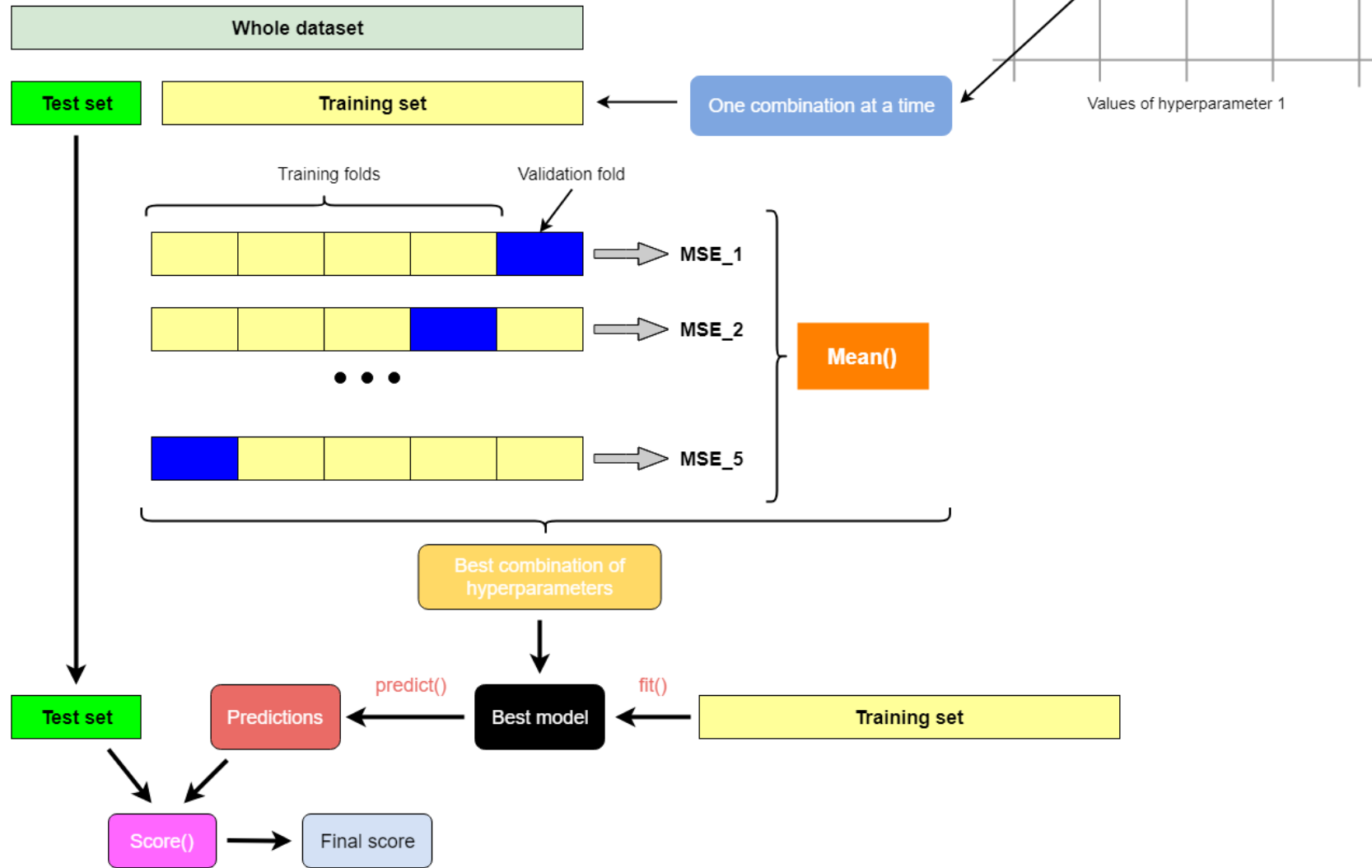


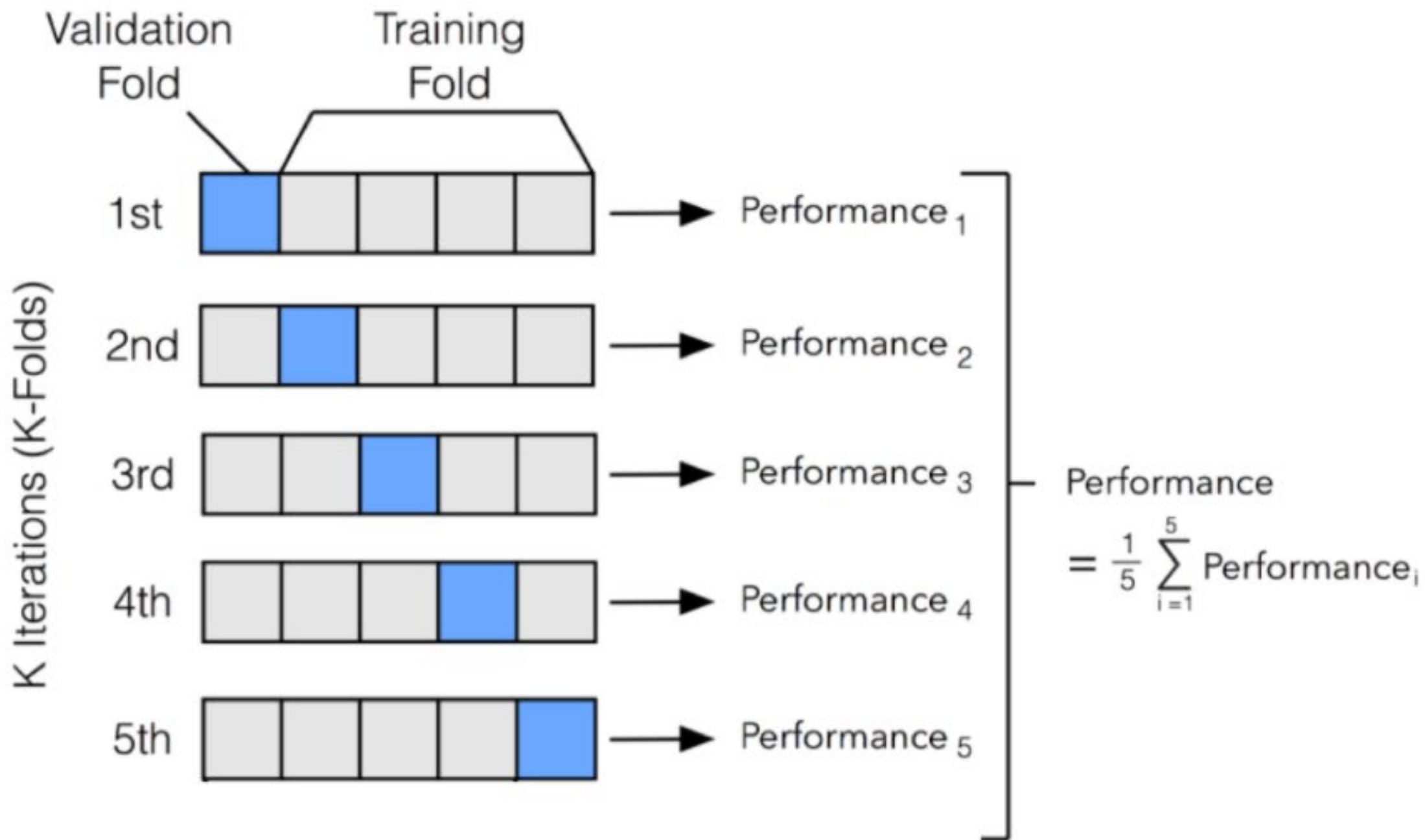
# CROSS VALIDATION

---

- Cross validation addresses this problem by repeating the split concept repeatedly.
- K-fold Cross Validation:
  - Randomly split the data into  $K$  subsets.
  - Use  $k-1$  set for training data.
  - Use the other set for testing data.
  - Repeat for all  $K$  subsets.
  - Average results together.
- Mitigates variation from data randomly in one set or the other.
- $K$  of 5 to 10 is typical.

## 5-fold cross-validation with grid search for hyperparameter tuning







# GRID SEARCH CV

---

- Grid Search CV combines these two into one simple call.
  - Test every HP combination.
  - Use cross validation to calculate scores.
  - Identify best model.
- In place of one “normal” score we get the cross-validated score, and best HP combo.
- Need to specify choices for HP in the grid.
  - Also a random version, that randomly chooses HP.
- A `gridsearchCV` is the “normal” tool to tune a model, at least for now.
  - We can try a model with many combos of HPs, and get the best one at the end.