UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

# Machine Learning

## Session 18 - T

## Ensemble Learning and Hyperparameter Optimization

**Degree in Applied Data Science**

**2024/2025**

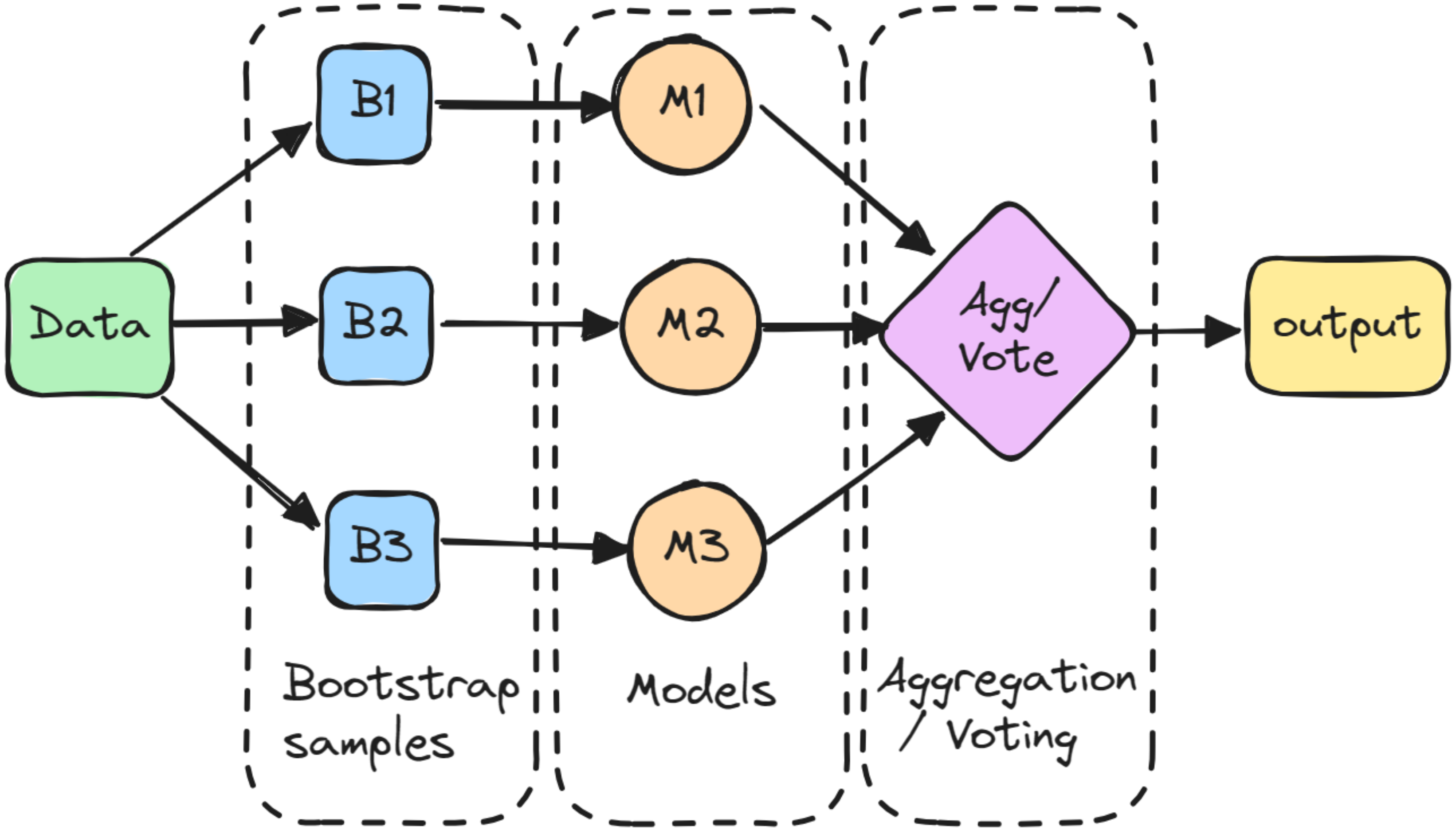
# Ensemble Learning

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Ensemble learning aims to build high-performance predictive models by **combining simpler base models**;

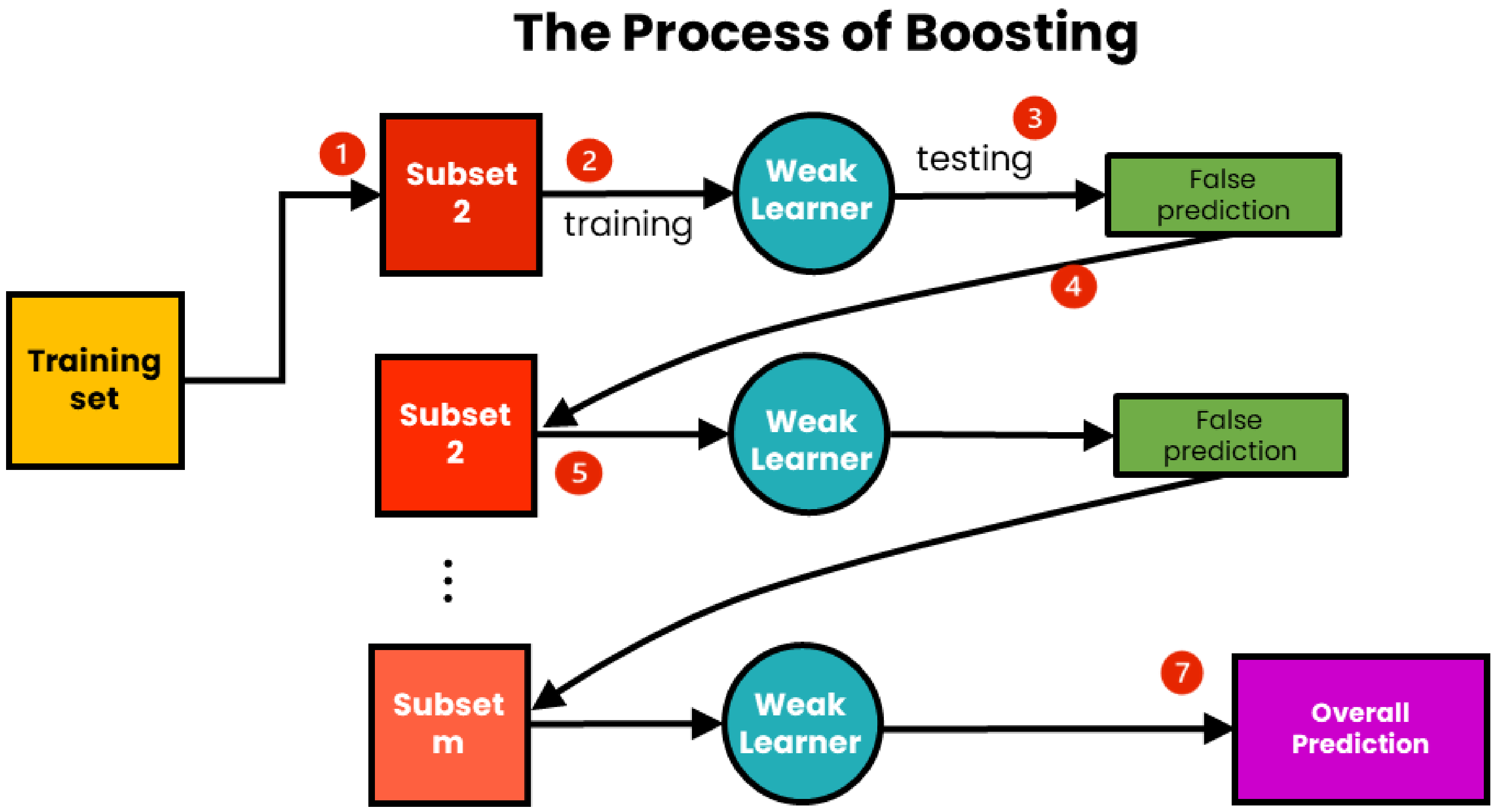
- There are different types of ensemble learning. The most common ones are:
  - **Bagging** (already covered in the Random Forests session!)
  - **Boosting** (already covered in the Random Forests session!)
  - **Stacking**
  - **Voting**

# Bagging Recap




https://www.datacamp.com/tutorial/what-bagging-in-machine-learning-a-guide-with-examples

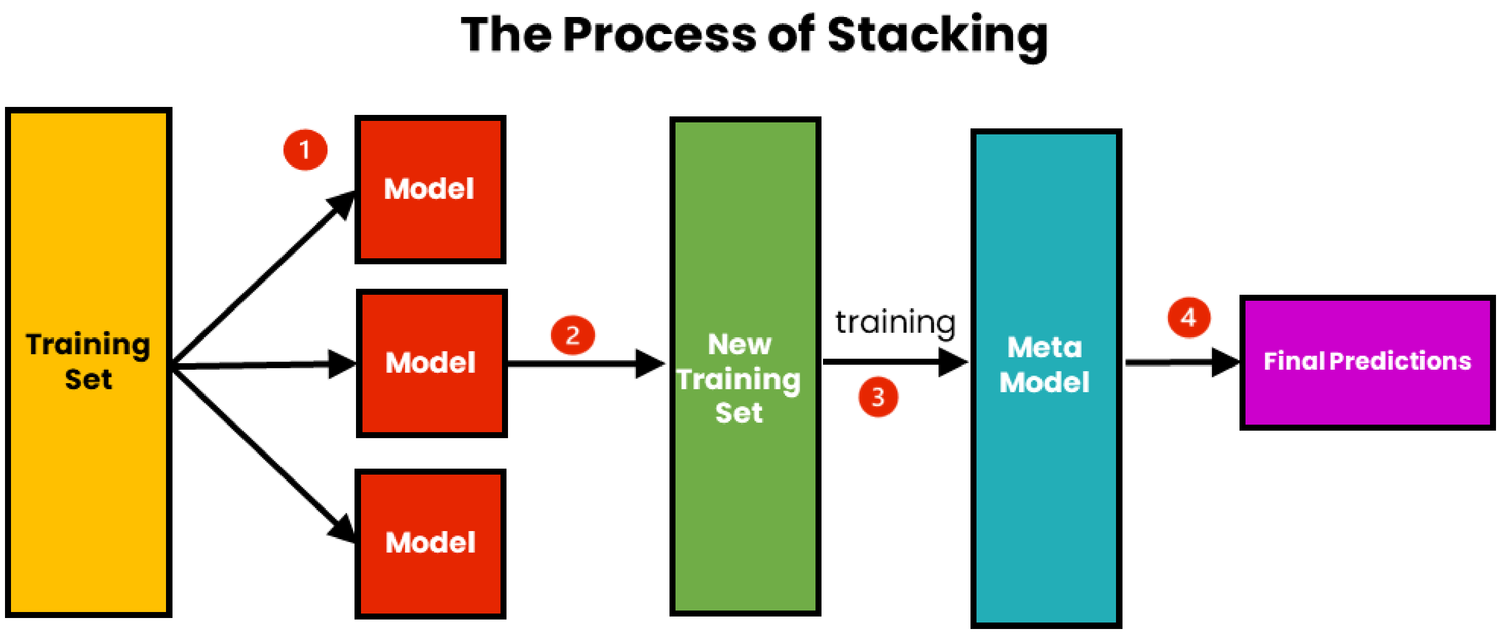
# Boosting Recap




https://medium.com/@brijesh_soni/understanding-boosting-in-machine-learning-a-comprehensive-guide-bdeaa1167a6

# Stacking

- The algorithm of stacking is:
  - **Train N models** on the training data;
  - **Train a meta model on the predictions** of the ensemble of models.




https://medium.com/@brijesh_soni/stacking-to-improve-model-performance-a-comprehensive-guide-on-ensemble-learning-in-python-9ed53c93ce28

# Bagging vs Boosting vs Stacking

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

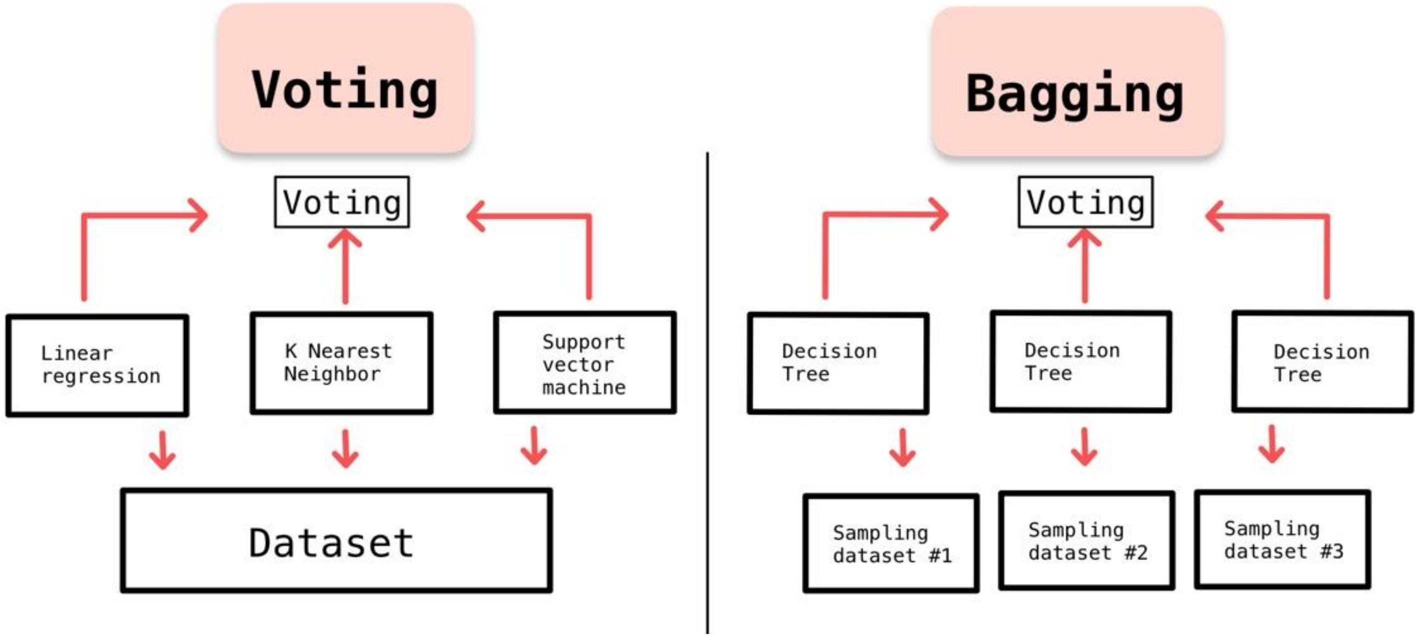
| | Bagging | Boosting | Stacking |
|---|---|---|---|
| Purpose | Reduce Variance | Reduce Bias | Improve Accuracy |
| Base Learner Types | Homogeneous | Homogeneous | Heterogeneous |
| Base Learner Training | Parallel | Sequential | Meta Model |
| Aggregation | Max Voting, Averaging | Weighted Averaging | Weighted Averaging |

https://www.analyticsvidhya.com/blog/2023/01/ensemble-learning-methods-bagging-boosting-and-stacking/

# Voting Ensemble

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

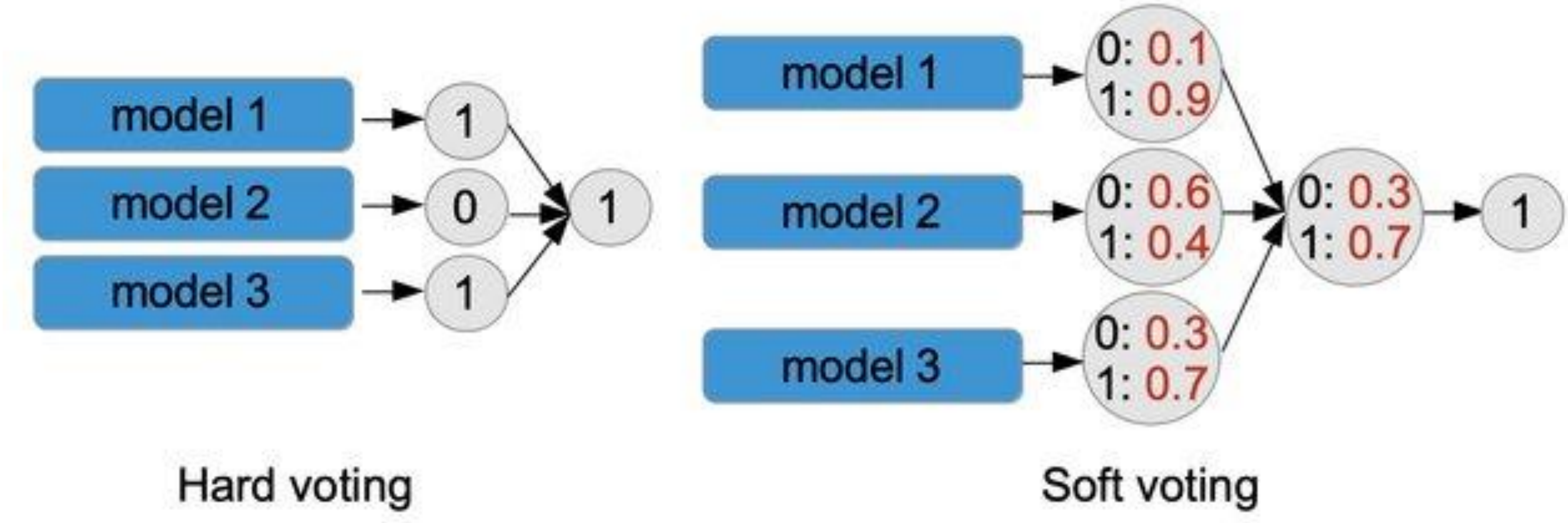
- Similar to Bagging without the bootstraping;
- Uses **heterogeneous models**;
- **Majory voting** for classification **averaging / weighted average** for regression.




https://medium.com/@chyun55555/ensemble-learning-voting-and-bagging-with-python-40de683b8ff0

# Voting Classifier

- **Hard vs Soft Voting**



# Hyperparameter Optimization

- Also called **metaparameter optimization**;
- Also called **hyperparameter tunig**;
- Any system that chooses hyperparameters automatically;
- **What is the difference between the model parameters and hyperparameters?**

# Hyperparameter Optimization

- What is a **hyperparameter**?
  - Parameters that the **user need to specify** in order to run a machine learning algorithm;

- What are some hyperparameters for the following models?
  - K-nearest-neighbot (kNN);
  - Decision Tree;
  - Neural Network.

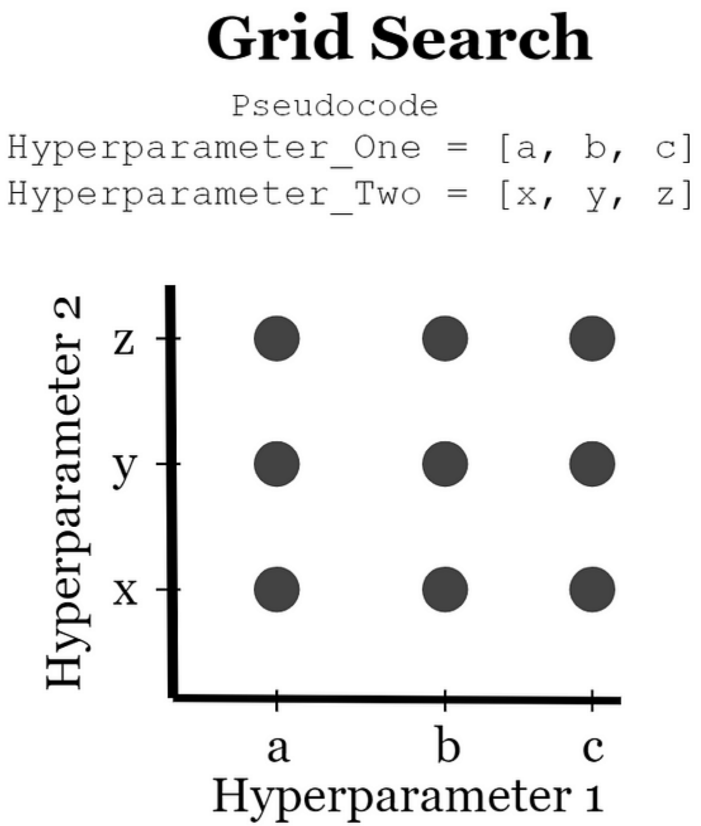
# Grid Search

- Define some **grid of parameters** you want to try;

- Try **all** the parameter values in the grid:
  - By **running the whole system** for each combination of parameters;

- Then **choose the combination with the best results**;

- It is essentially a **brute force method**.



# Grid Search - Disadvantages

- As the number of parameters increases, the cost of grid search **increases exponentially**!
  - Why?

- Still need some way to choose the grid properly:
  - This can be as hard as the original hyperparameter optimization;

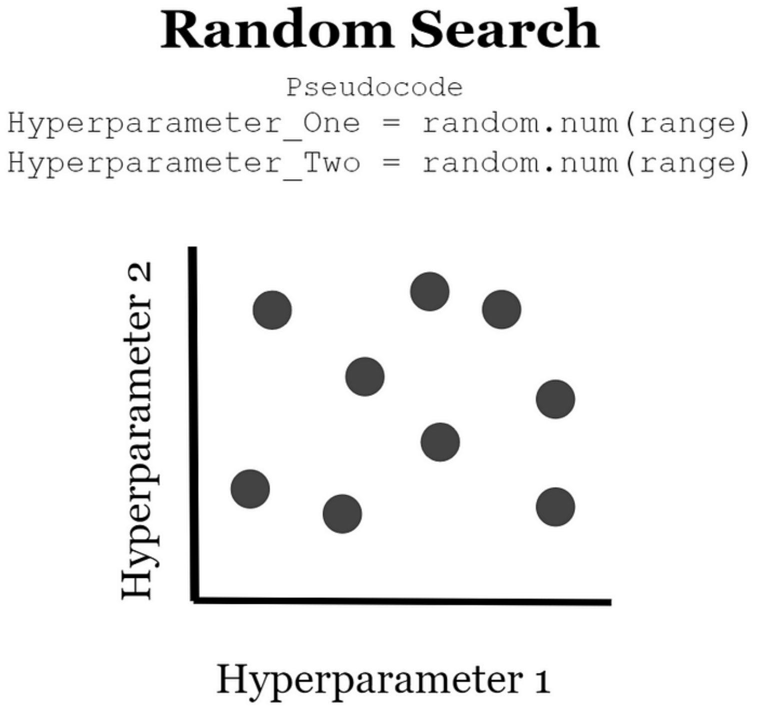
- Can't take advantage of any insight you have about the system!

# Randomized Search

- This is similar to grid search, but instead of trying all hyperparameter combinations try some **random combinations**;

- This solves the curse of dimentionality:
  - The number of trials do not grow exponetially as the number odf dimentions increases;

- Problem: we do **not necessarily get anywhere near the optimal hyperparameters** in a small sample.



# Bayesian Hyperparameter Optimization

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Bayesian approaches, in contrast to random or grid search, **keep track of past evaluation results**;

- They form a **probabilistic model** mapping hyperparameters to a probability of a score on the objective function:
  - P(score | hyperparameters)

- This is called a "surrogate" for the objective function. This surrogate is **much easier to optimize** than the objective function;

# Bayesian Hyperparameter Optimization

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Bayesian optimization steps:

  1. Build a surrogate probability model of the objective function

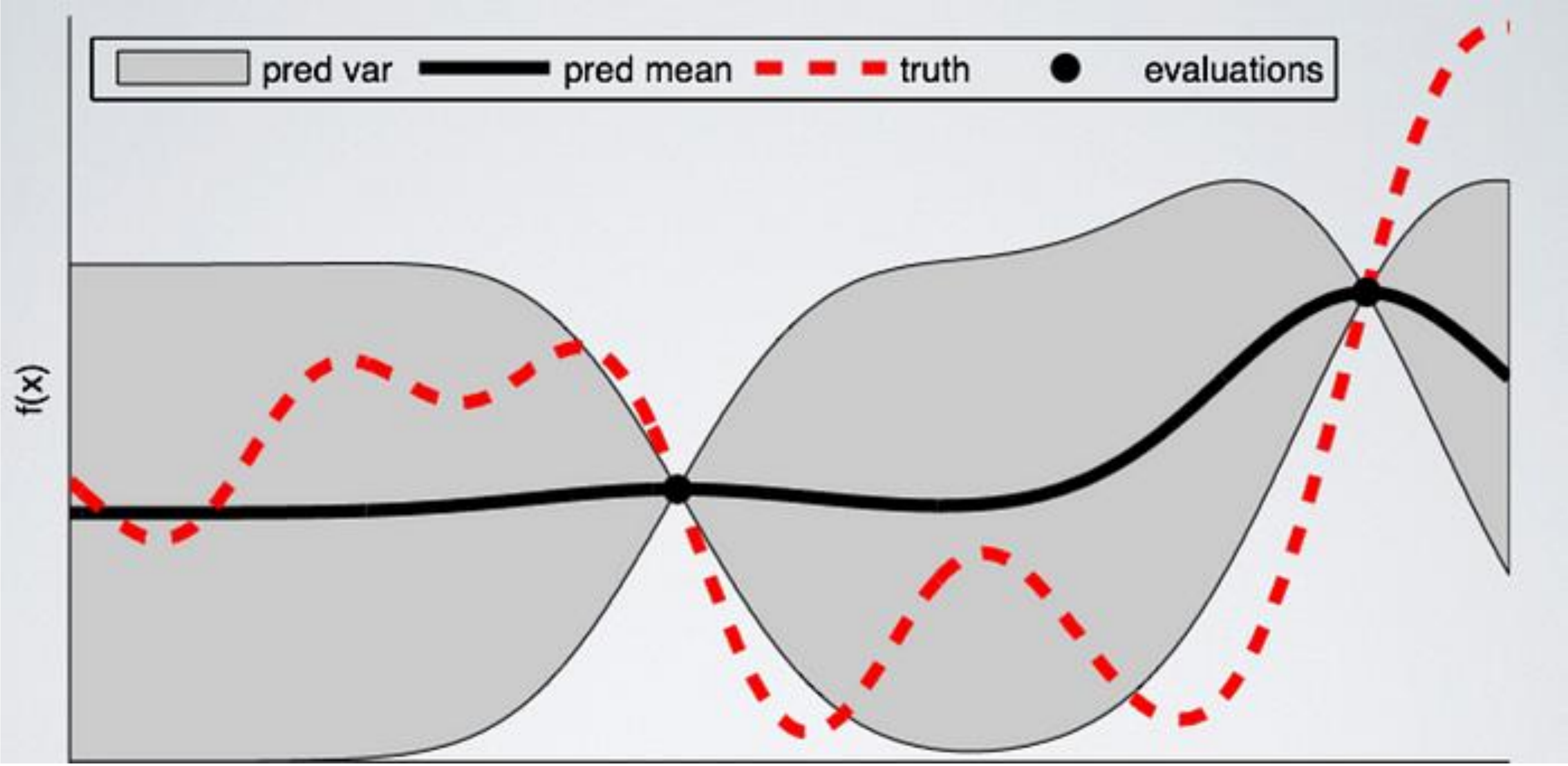
  2. Find the hyperparameters that perform best on the surrogate

  3. Apply these hyperparameters to the true objective function

  4. Update the surrogate model incorporating the new results

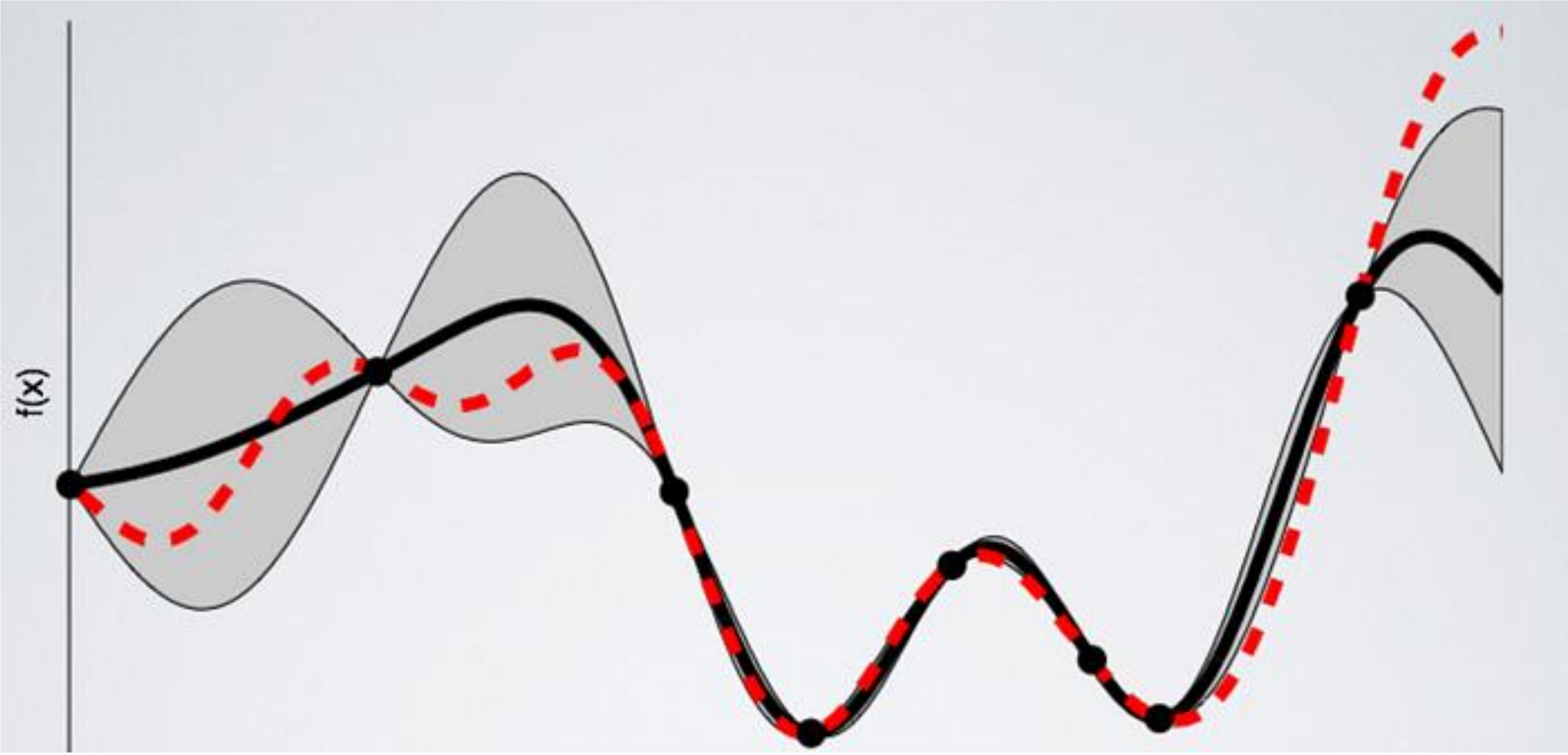
  5. Repeat steps 2–4 until max iterations or time is reached

# Bayesian Hyperparameter Optimization



# Bayesian Hyperparameter Optimization

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA



# Resources


- Rokach, L. (2019). Ensemble Learning Pattern Classification using ensemble methods. World Scientific.

- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. In WIREs Data Mining and Knowledge Discovery (Vol. 13, Issue 2). Wiley. https://doi.org/10.1002/widm.1484