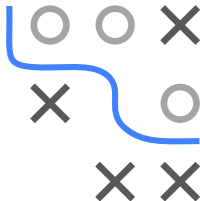# Introduction to Machine Learning

# Tuning: In a Nutshell



**Learning goals**

- Understand the main idea behind tuning,
- why tuning matters,
- and why tuning is difficult

# WHAT IS TUNING?

- Tuning is the process of selecting the best hyperparameters, denoted as $\lambda$, for a machine learning model
- Hyperparameters are the parameters of the learner (versus model parameters $\theta$)
- Consider a guitar analogy: Hyperparameters are akin to the tuning pegs. Learning the best parameters $\hat{\theta}$ - playing the guitar - is a separate process that depends on tuning!
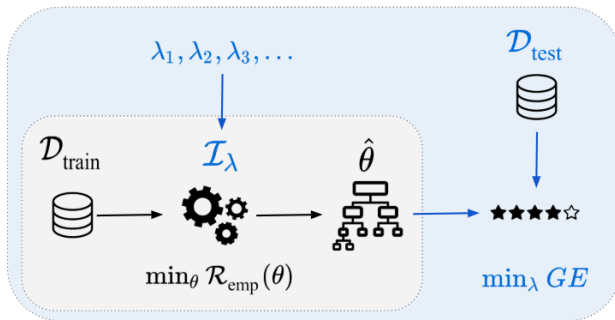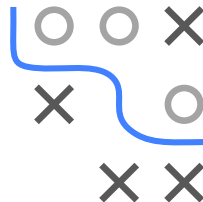
# WHY TUNING MATTERS

- Just like a guitar won't perform well when out-of-tune, properly tuning a learner can drastically improve the resulting model performance
- Tuning helps find a balance between underfitting and overfitting
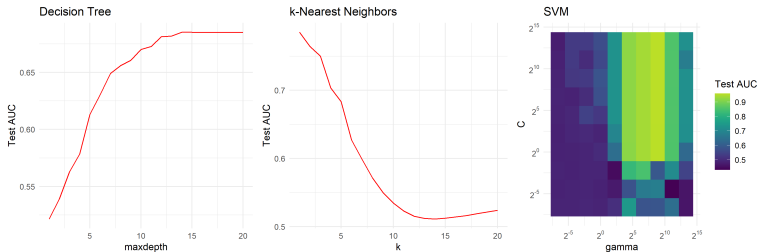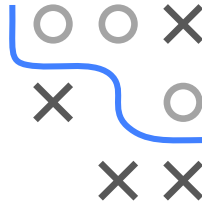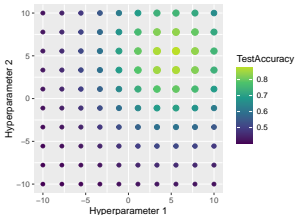
# HOW HARD COULD IT BE?

- Very difficult: There are lots of different configurations to choose from, known as the hyperparameter space, denoted by $\Lambda$ (analogous to $\Theta$)
- Black box: If one opts for a configuration $\lambda \in \Lambda$, how can its performance be measured (and compared)?
- Well-thought-out approaches - black box optimization techniques - are needed!
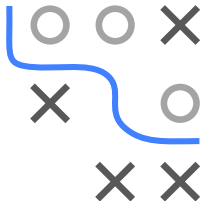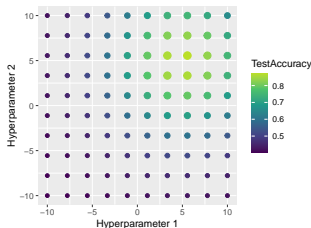
# NAÏVE APPROACHES

Let's start with two naïve approaches -
**Grid Search** and **Random Search**:
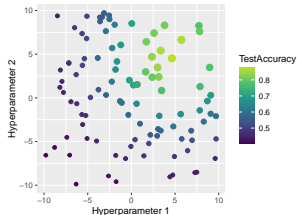


Beyond these basic methods, there are more sophisticated techniques which
operate on certain assumptions about the objective function. These
assumptions enable them to search for optimal solutions more efficiently.
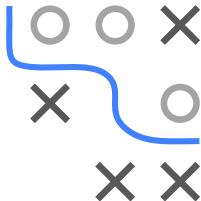
# PIPELINES IN MACHINE LEARNING

Pipelines are like the assembly lines in machine learning. They automate the sequence of data processing and model building tasks, ensuring efficiency and consistency.

**Why Pipelines Matter:**

- **Streamlined Workflow:** Automates the flow from data preprocessing to model training.
- **Reproducibility:** Ensures that results can be reproduced consistently.
- **Error Reduction:** Minimizes the chance of human errors in the model building process.

**Simple Pipeline Example:**

- A basic pipeline might include data normalization, feature selection, and a learning algorithm.

# PIPELINES AND AUTOML

AutoML leverages pipelines to automate the process of applying machine learning to real-world problems. It simplifies tasks like model selection, hyperparameter tuning, and cross-validation.

**Key Components of AutoML Pipelines:**

- **Data Preprocessing:** Automatic handling of missing values, encoding categorical variables, etc.
- **Feature Engineering:** Automated feature selection and transformation.
- **Model Selection:** Evaluating and choosing the best model automatically.
- **Hyperparameter Optimization:** Finding the best model settings without manual intervention.