

Guided Tour of Machine Learning in Finance

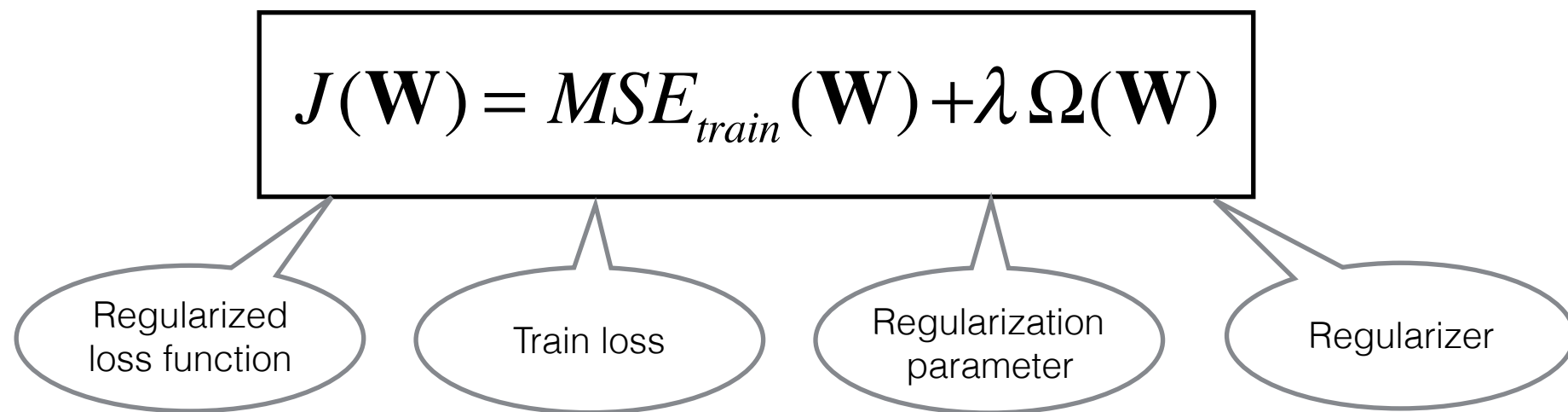
Regularization, validation set, and hyper-parameters

Igor Halperin

NYU Tandon School of Engineering, 2017

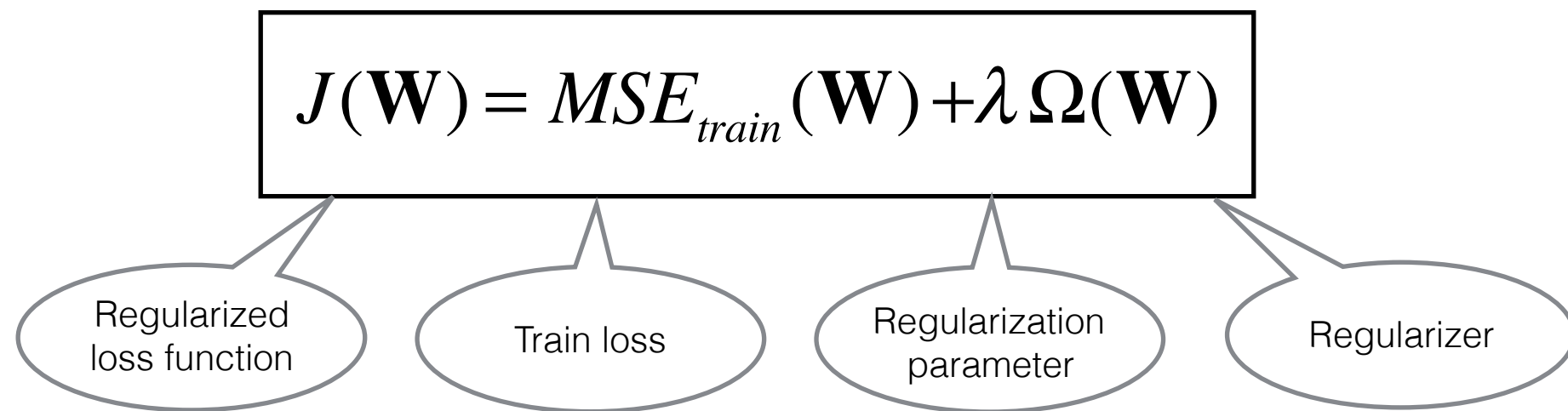
Regularization

Recall that we minimize MSE_{train} , though want in fact to minimize MSE_{test}
The idea of regularization is to modify the objective function of minimization of MSE_{train} so that MSE_{test} has a smaller variance:



Regularization

Recall that we minimize MSE_{train} , though want in fact to minimize MSE_{test} . The idea of regularization is to modify the objective function of minimization of MSE_{train} so that MSE_{test} has a smaller variance:



Popular choices for the **regularizer**:

- L_2 regularization $\Omega(\mathbf{W}) = \mathbf{W}^T \mathbf{W} = \|\mathbf{W}\|_2$

Penalizes large weights

- L_1 regularization $\Omega(\mathbf{W}) = \sum_i |W_i| = \|\mathbf{W}\|_1$

Enforces a sparse solution

- Entropy regularization $\Omega(\mathbf{W}) = \sum_i W_i \log W_i$ ($W_i \geq 0, \sum_i W_i = 1$)

Motivated by Bayesian statistics

Hyperparameters and validation set

- **Hyperparameters** are any quantitative features of ML models that are not directly optimized by minimizing in-sample loss such as $\text{MSE}_{\text{train}}$
- **Hyperparameters control model capacity**

Hyperparameters and validation set

- **Hyperparameters** are any quantitative features of ML models that are not directly optimized by minimizing in-sample loss such as $\text{MSE}_{\text{train}}$
- **Hyperparameters control model capacity**
- Examples of hyperparameters:
 - ◆ Degree of a polynomial regression (linear, quadratic, cubic, etc.)
 - ◆ Regularization parameter λ
 - ◆ Number of levels in a decision tree
 - ◆ Number of layers and nodes per layer in neural networks
 - ◆ Learning rate
 - ◆ ...

Hyperparameters and validation set

- **Hyperparameters** are any quantitative features of ML models that are not directly optimized by minimizing in-sample loss such as $\text{MSE}_{\text{train}}$
- **Hyperparameters control model capacity**
- Examples of hyperparameters:
 - ✦ Degree of a polynomial regression (linear, quadratic, cubic, etc.)
 - ✦ Regularization parameter λ
 - ✦ Number of levels in a decision tree
 - ✦ Number of layers and nodes per layer in neural networks
 - ✦ Learning rate
 - ✦ ...
- How to choose hyperparameters:
 - ✦ Split a training set into training and **validation** sets (e.g. as 80:20)

Hyperparameters and validation set

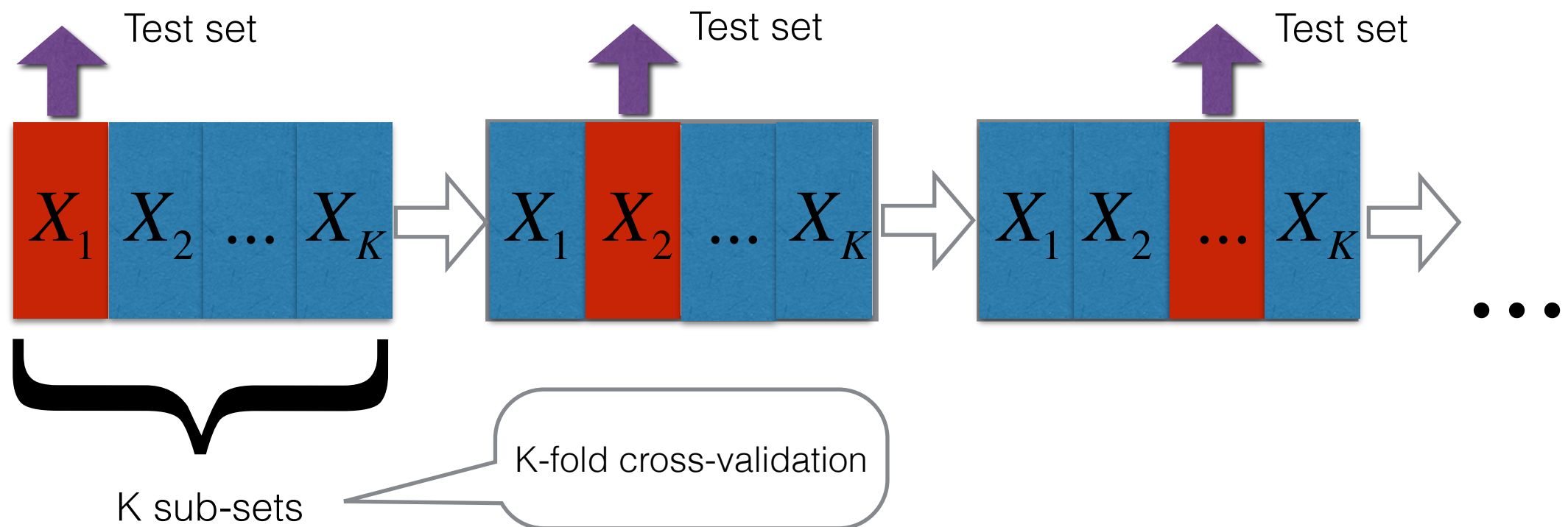
- **Hyperparameters** are any quantitative features of ML models that are not directly optimized by minimizing in-sample loss such as $\text{MSE}_{\text{train}}$
- **Hyperparameters control model capacity**
- Examples of hyperparameters:
 - ◆ Degree of a polynomial regression (linear, quadratic, cubic, etc.)
 - ◆ Regularization parameter λ
 - ◆ Number of levels in a decision tree
 - ◆ Number of layers and nodes per layer in neural networks
 - ◆ Learning rate
 - ◆ ...
- How to choose hyperparameters:
 - ◆ Split a training set into training and **validation** sets (e.g. as 80:20)
 - ◆ Use **cross-validation**

Cross-validation

- Assume we are given N samples, but N is small, so setting aside a fixed test set is problematic.
- We want to use all samples for training!
- This is achieved using cross-validation:

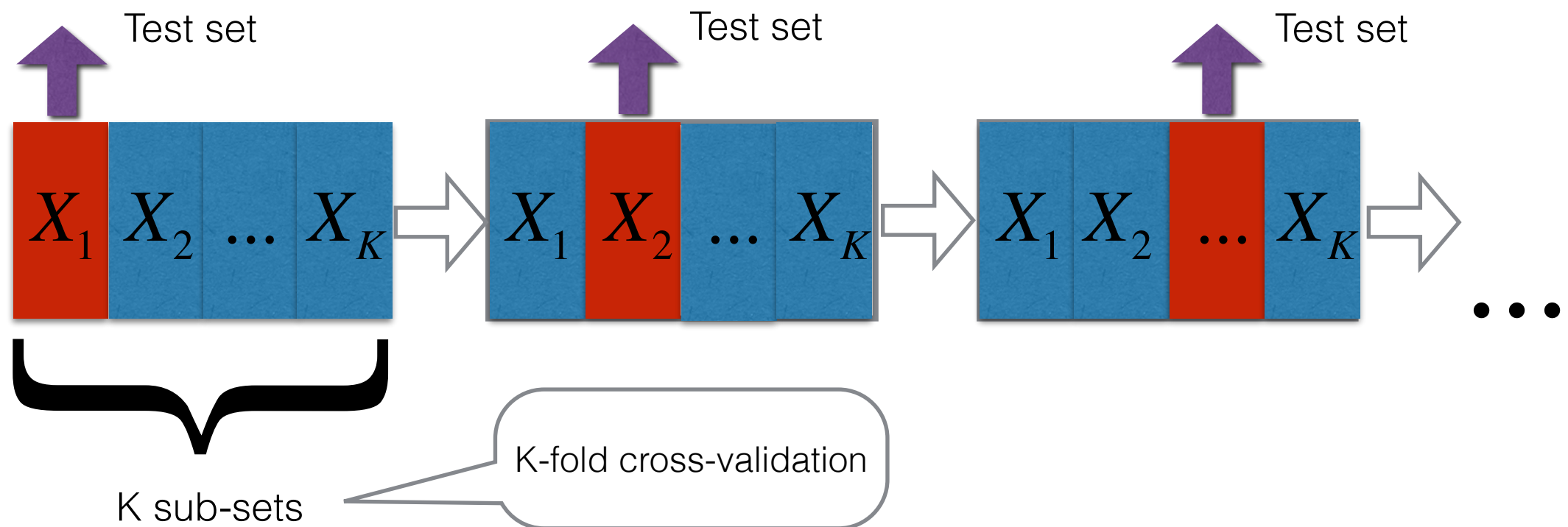
Cross-validation

- Assume we are given N samples, but N is small, so setting aside a fixed test set is problematic.
- We want to use all samples for training!
- This is achieved using cross-validation:



Cross-validation

- Assume we are given N samples, but N is small, so setting aside a fixed test set is problematic.
- We want to use all samples for training!
- This is achieved using cross-validation:



Special cases:

- $K=1$: no test sub-set!
- $K=N$: leave-on-out cross-validation

Summary: Supervised Learning diagram

General diagram for training supervised learning algorithms

