Aprendizado de Máquina e Reconhecimento de Padrões 2021.2

Hyperparameter Optimization (Fine-tuning)

Prof. Dr. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas

Model Hyperparameters

Properties that are **external** to the model and whose value **cannot be estimated/learned from data**.

Examples:

- **Imputer's strategy**: 'median'
- **Number of neighbors for KNN:** 3

Model Parameters

Properties that are **internal** to the model and whose value **can be estimated/learned from data**.

Examples:

- **Estimated value for missing values:** 20 (median)
- **Estimated coefficients of a linear regression.**

Problem

Filling in missing values (imputer)

	Name	Age
0	John	19.0
1	Maria	22.0
2	Alice	18.0
3	Margot	20.0
4	Pedro	NaN
5	Giovanni	21.0

(training) data

(hyperparameter)

Imputer's strategy: 'median'

(parameter)

estimated/learned value: 20

Hyperparameter Optimization (Fine-tuning)

- It is the problem of choosing a set of **optimal values** **for** **hyperparameters** for a **learning algorithm** and **data**.

hyperparameters Ψ

A	B
1	9
3	14
$\boxed{4}$	$\boxed{12}$
8	18
12	20

$$\boldsymbol{\varphi}^* = \text{argmax}_{\boldsymbol{\varphi}_i \epsilon \Psi} f(\boldsymbol{\varphi})$$

Or **argmin** for **error functions**.

Ψ: all hyperparameter combinations/sets

$\boldsymbol{\varphi}_i$: i-th hyperparameter combination/set from Ψ

$f(\boldsymbol{\varphi})$: training and validation of the ML algorithm with $\boldsymbol{\varphi}$

$\boldsymbol{\varphi}^*$: optimum hyperparameter combination

Holdout Strategy

1 Split your data into train, validation, and test sets

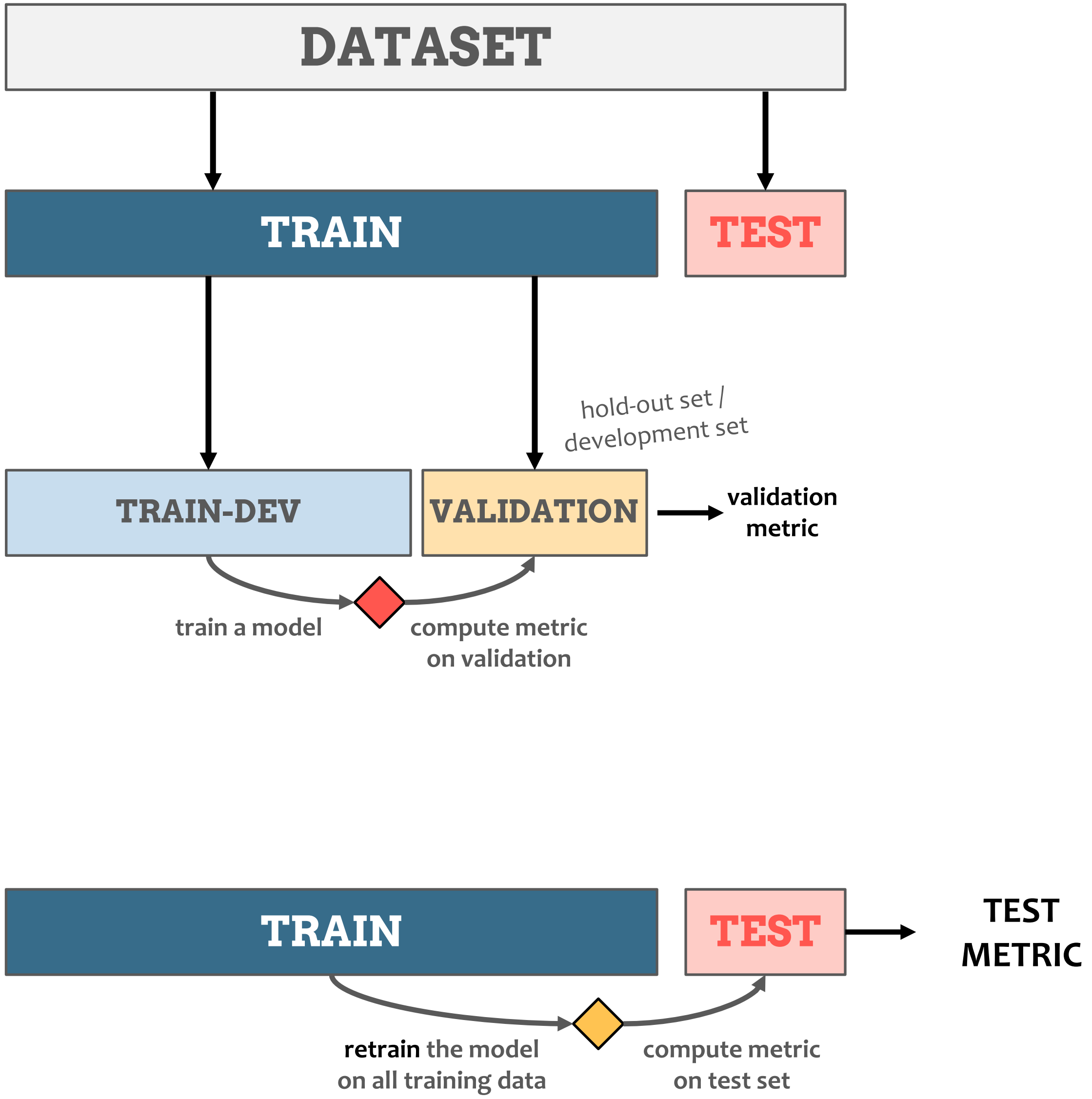
2 For each hyperparameter combination $\varphi_i \in \Psi$

A	B
1	9
3	14
4	12
8	18
12	20

hyperparameters

3 Choose the hyperparameter combination with the best metric

φ^*: A | 1 | 18 | B



Holdout Strategy

1 Split your data into train, validation, and test sets

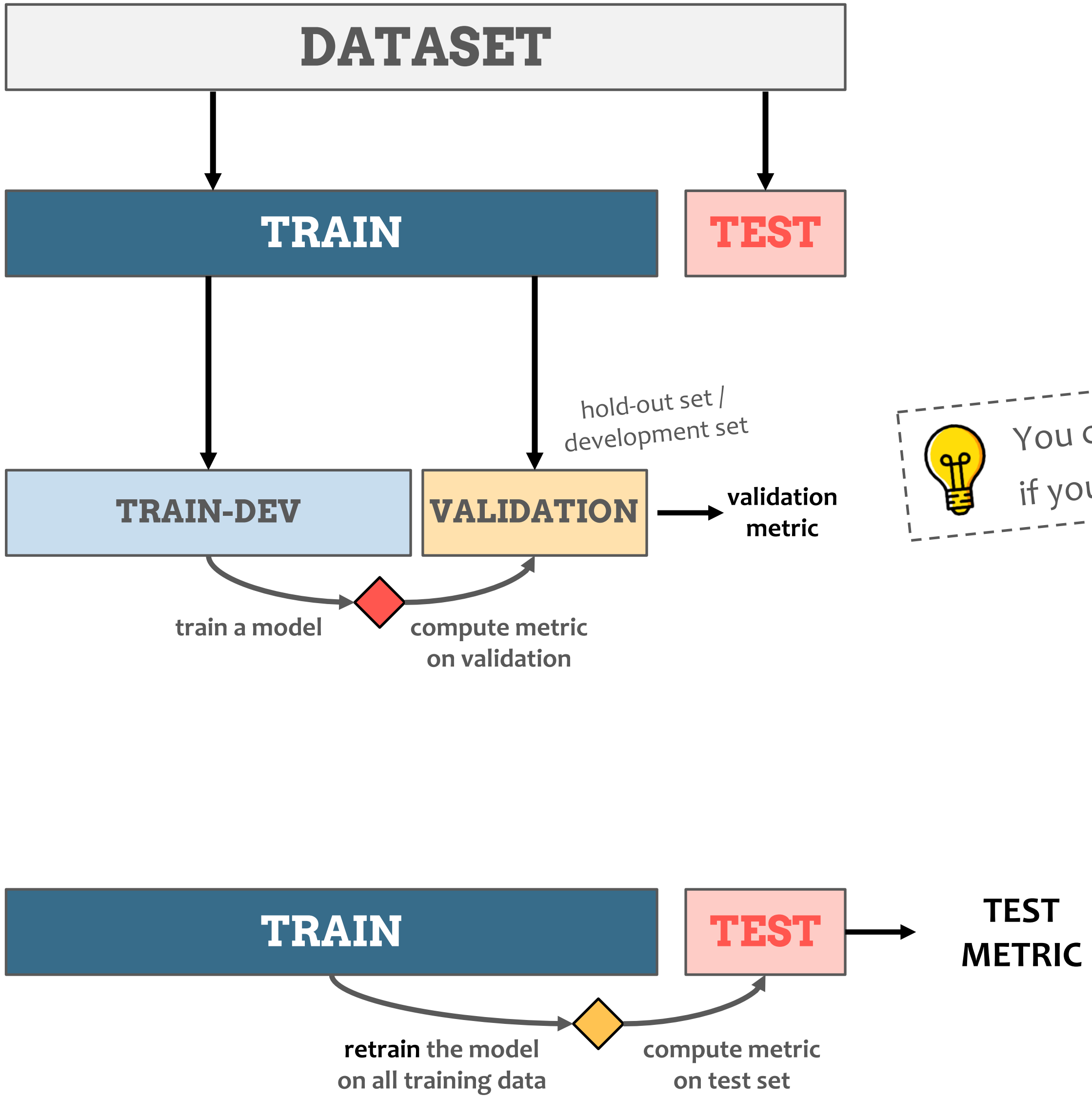
2 For each hyperparameter combination $\varphi_i \epsilon \Psi$

A	B
1	9
3	14
4	12
8	18
12	20

hyperparameters

3 Choose the hyperparameter combination with the best metric

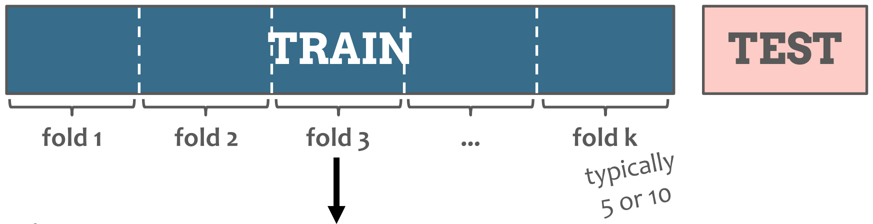
φ^*: **A** 1 18 **B**



You can use **stratified sampling** if you want.

k-Fold Strategy

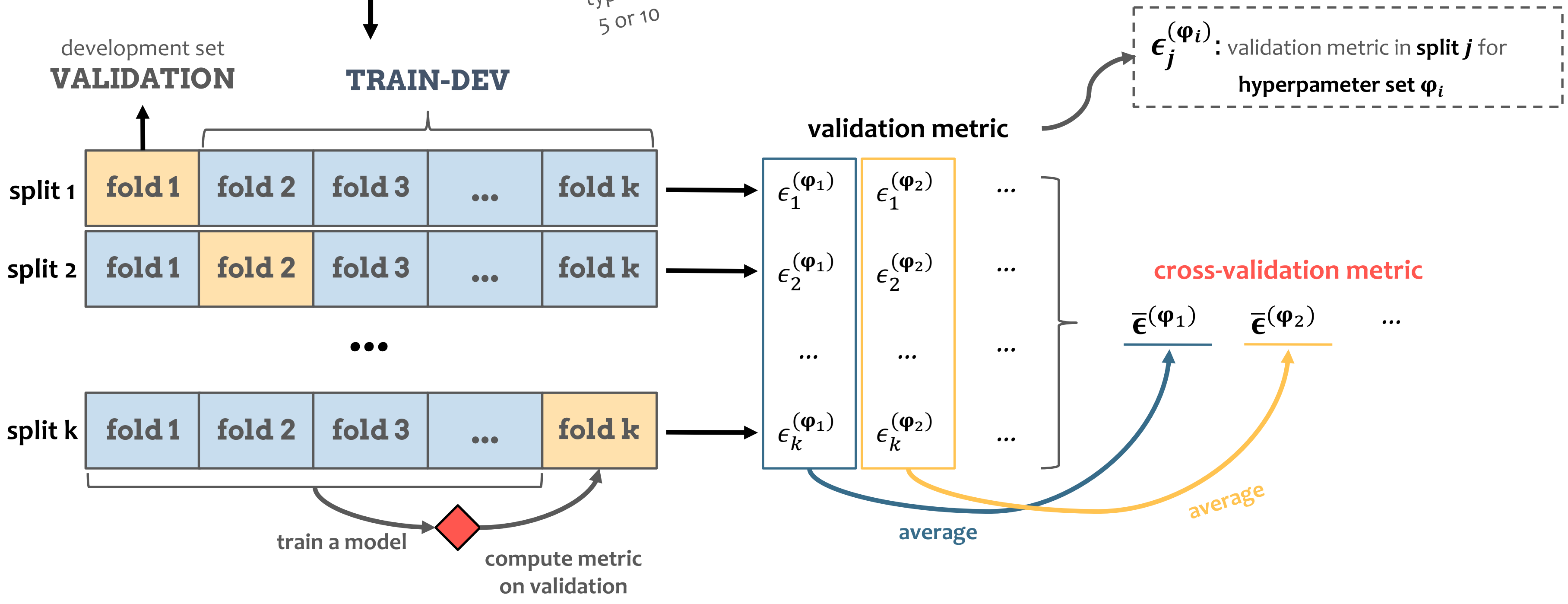
1 Set aside the test set and split the train set into k folds



2 For each hyperparameter combination $\varphi_i \epsilon \Psi$

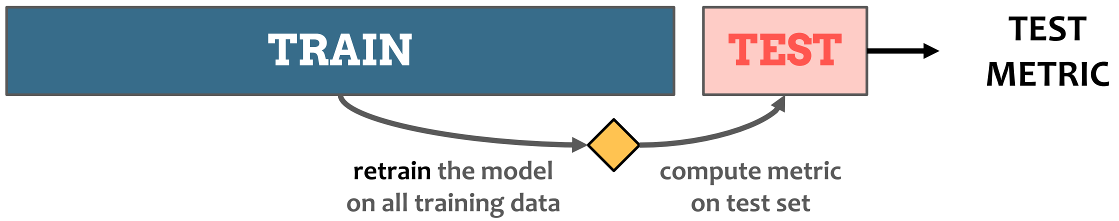
A	B
1	9
3	14
4	12
8	18
12	20

hyperparameters



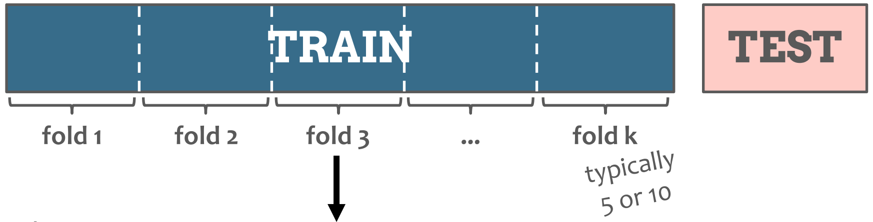
3 Choose the hyperparameter combination with the best metric

φ^*: A 1 18 B



k-Fold Strategy

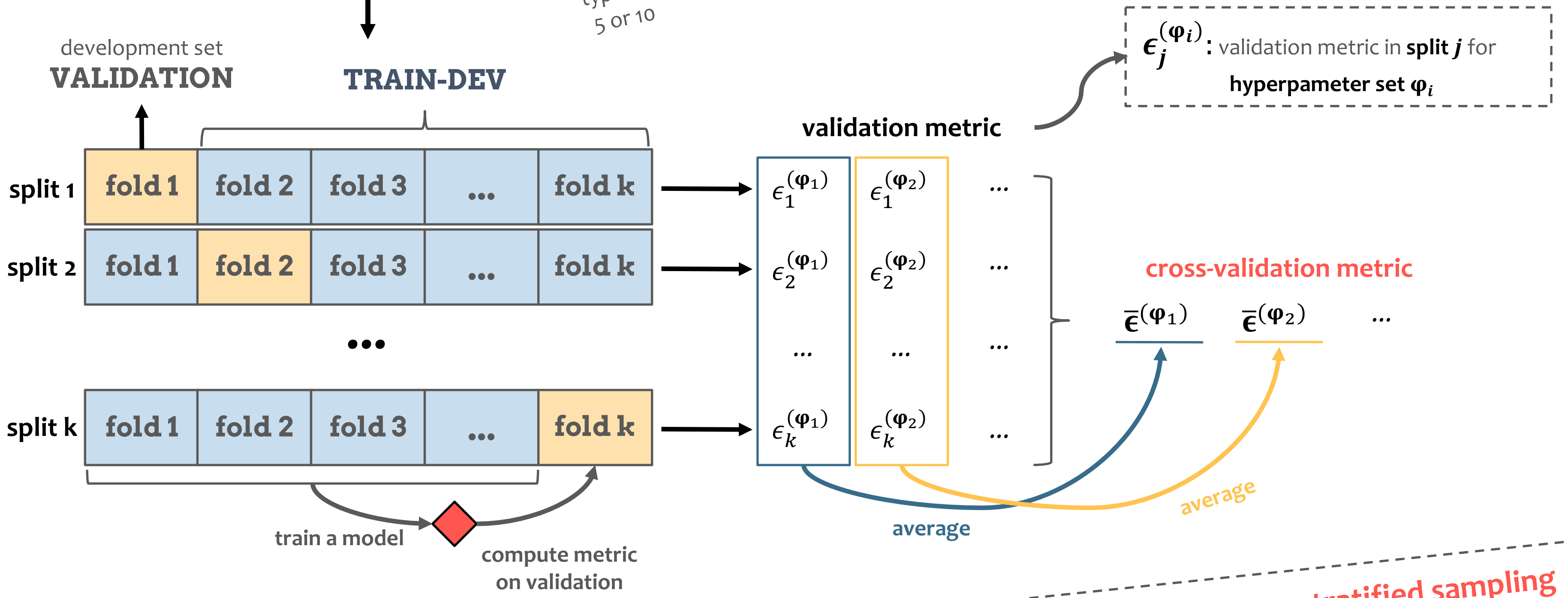
1 Set aside the test set and split the train set into k folds



2 For each hyperparameter combination $\varphi_i \in \Psi$

A	B
1	9
3	14
4	12
8	18
12	20

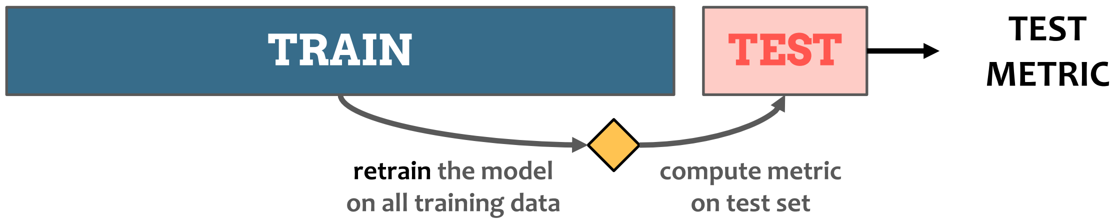
hyperparameters



You can use **stratified sampling** if you want.

3 Choose the hyperparameter combination with the best metric

φ^*: A 1 18 B



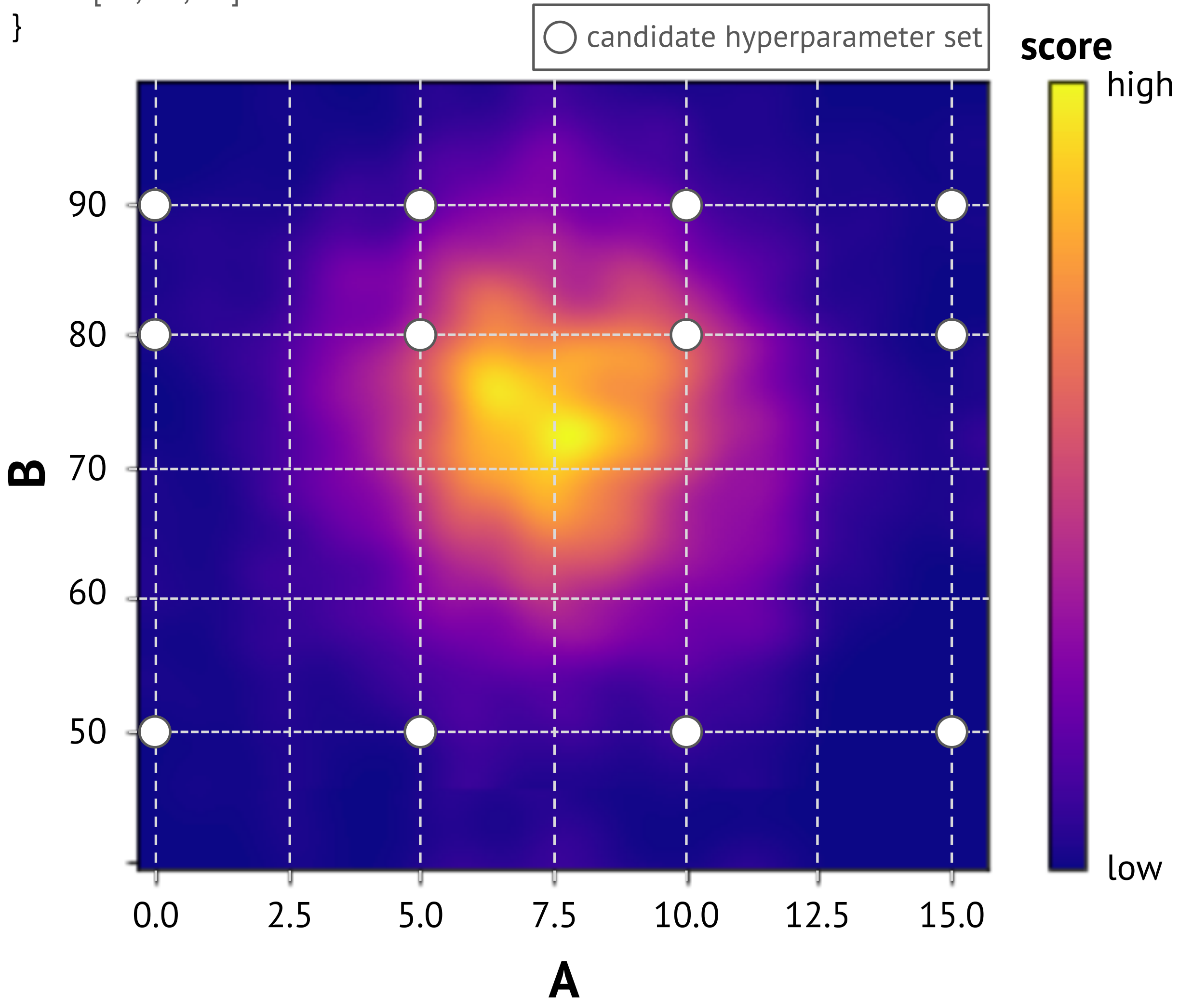
Search Space for Fine-Tuning

Grid Search

`sklearn.model_selection.GridSearchCV`

hyperparameters

search space = {
  '**A**': [0.0, 5.0, 10.0, 15.0],
  '**B**': [50, 80, 90]
}



The **search space may not** include the **optimum hyperparameter combination**: the highest score (lowest error).

The **larger** the **search space**, the **longer (exponentially)** the grid search.

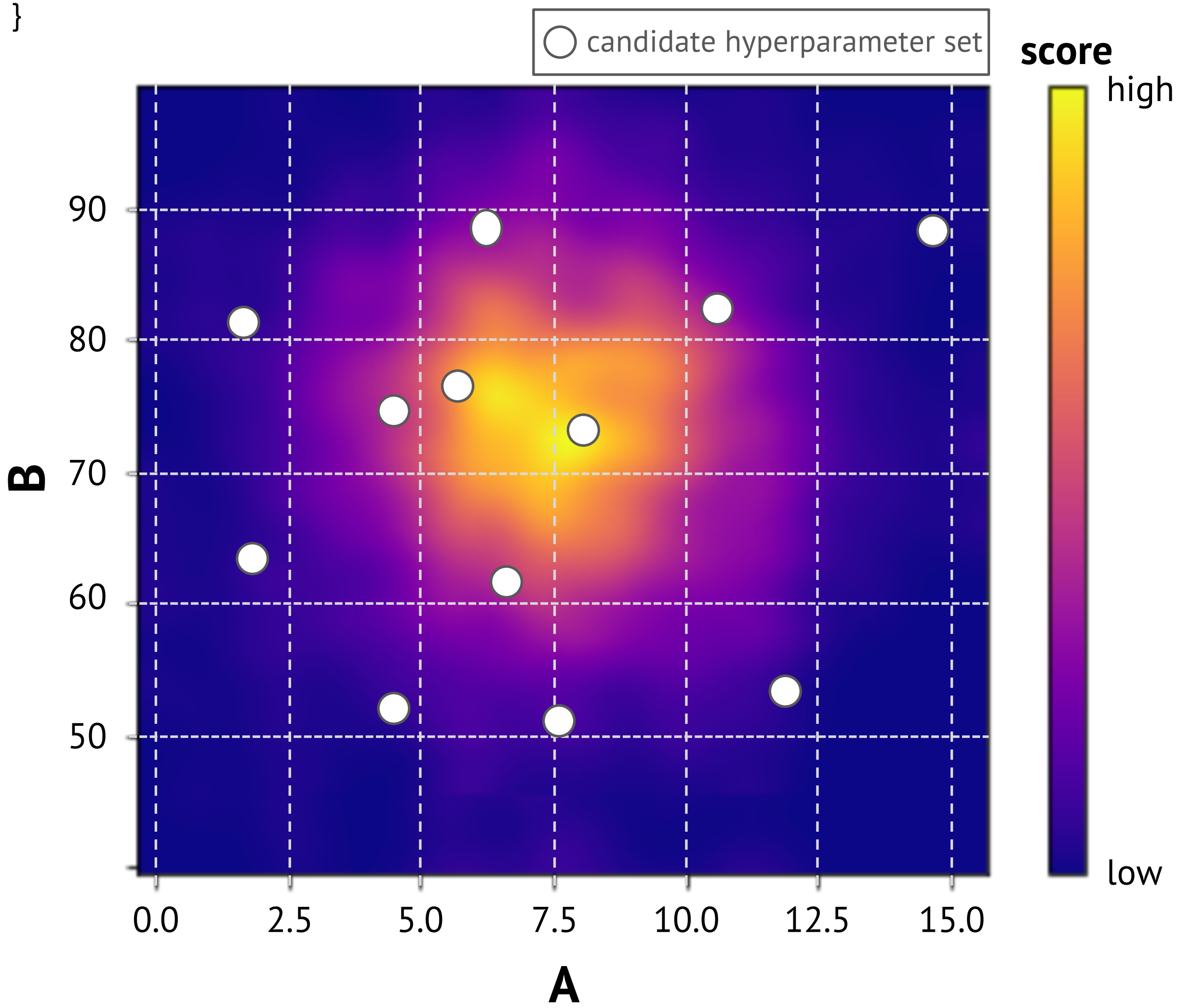
Randomized Search

n = 9

sklearn.model_selection.RandomizedSearchCV

hyperparameters

search space = {
'A': [0.0, 0.25, 0.5, ..., 14.75, 15.0],
'B': [50, 51, 52, ..., 86, 87, 88, 89, 90]
}



You can even increase the **hyperparameter distribution** for fine-tunning and keep **a feasible processing time.**

Aprendizado de Máquina e Reconhecimento de Padrões 2021.2

Hyperparameter Optimization (Fine-tuning)

Prof. Dr. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas