# What we interesting when evaluate an algorithm

In every step as possible:
how the training data was prepared (e.g. scaling),
the choice of algorithm (e.g. kNN),
and how the chosen algorithm was configured (e.g. k=3).

We generalize the performance measure from:

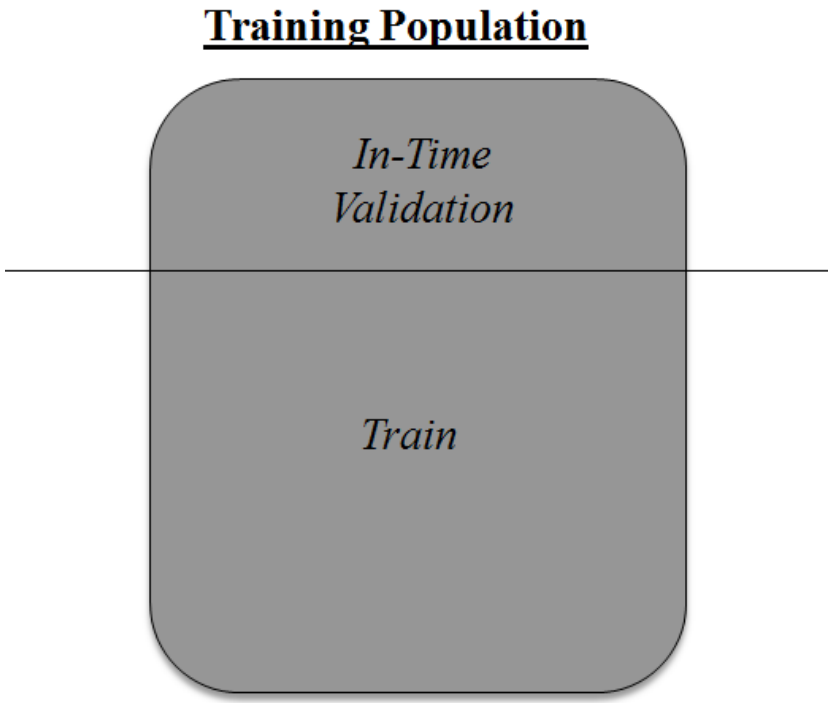
- "*the skill of the procedure on the **test set***"

to

- "*the skill of the procedure on **unseen data***".

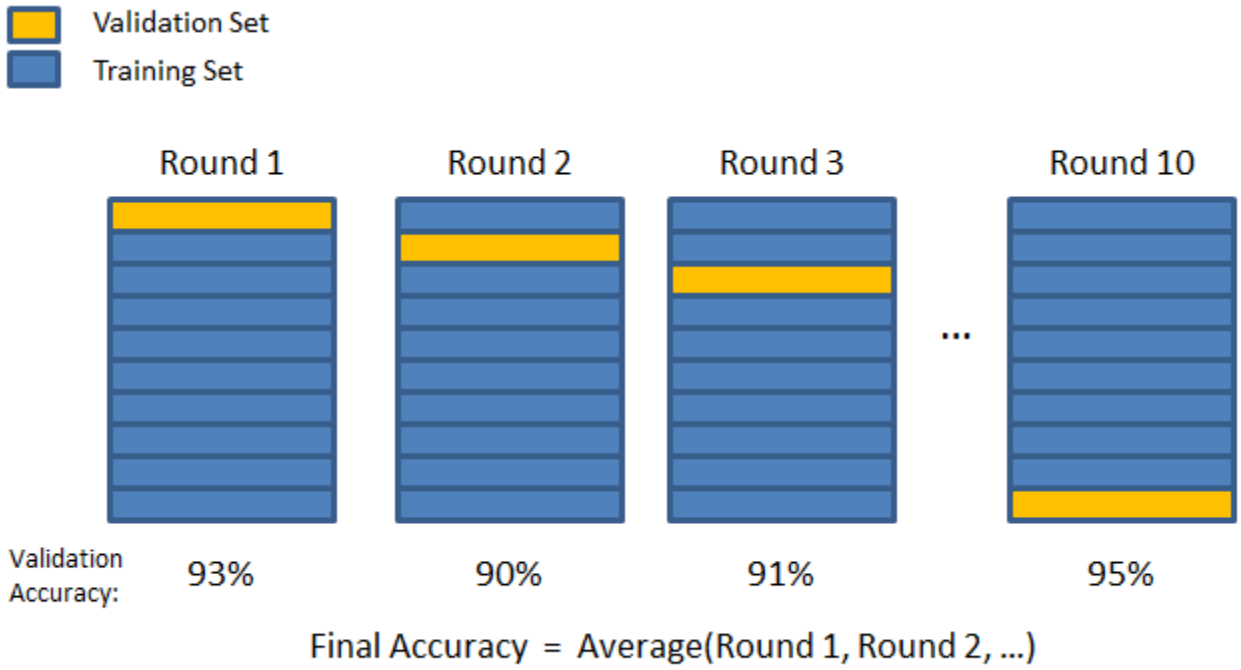
INVESTIC

# Cross Validation

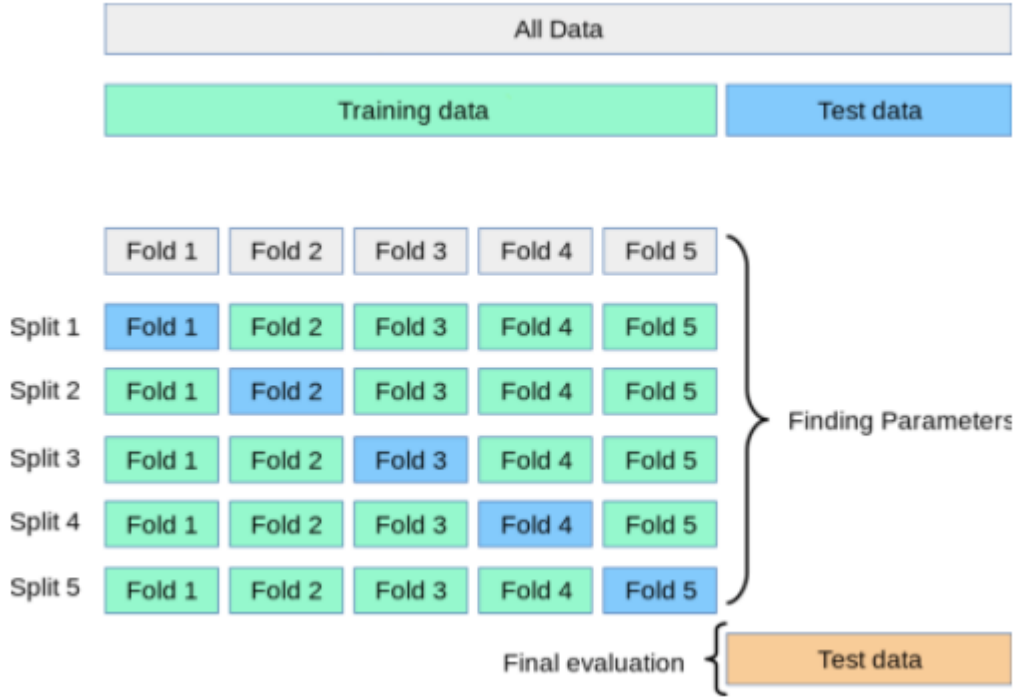



INVESTIC

# K Fold Cross Validation



# Good Practice??



# The Purpose of k-fold Cross Validation

Estimate the skill of a method on unseen data. Like using a train-test split.

Cross-validation systematically creates and evaluates multiple models on multiple subsets of the dataset. **This, in turn, provides a population of performance measures**.

- We can calculate the mean of these measures to get an idea of **how well the procedure performs on average.**
- We can calculate the standard deviation of these measures to get an idea of **how much the skill of the procedure is expected to vary in practice.**

# Then?

## How to choose a predictive model after k-fold cross-validation?

Asked 8 years, 7 months ago  Active 16 days ago  Viewed 159k times

228

260

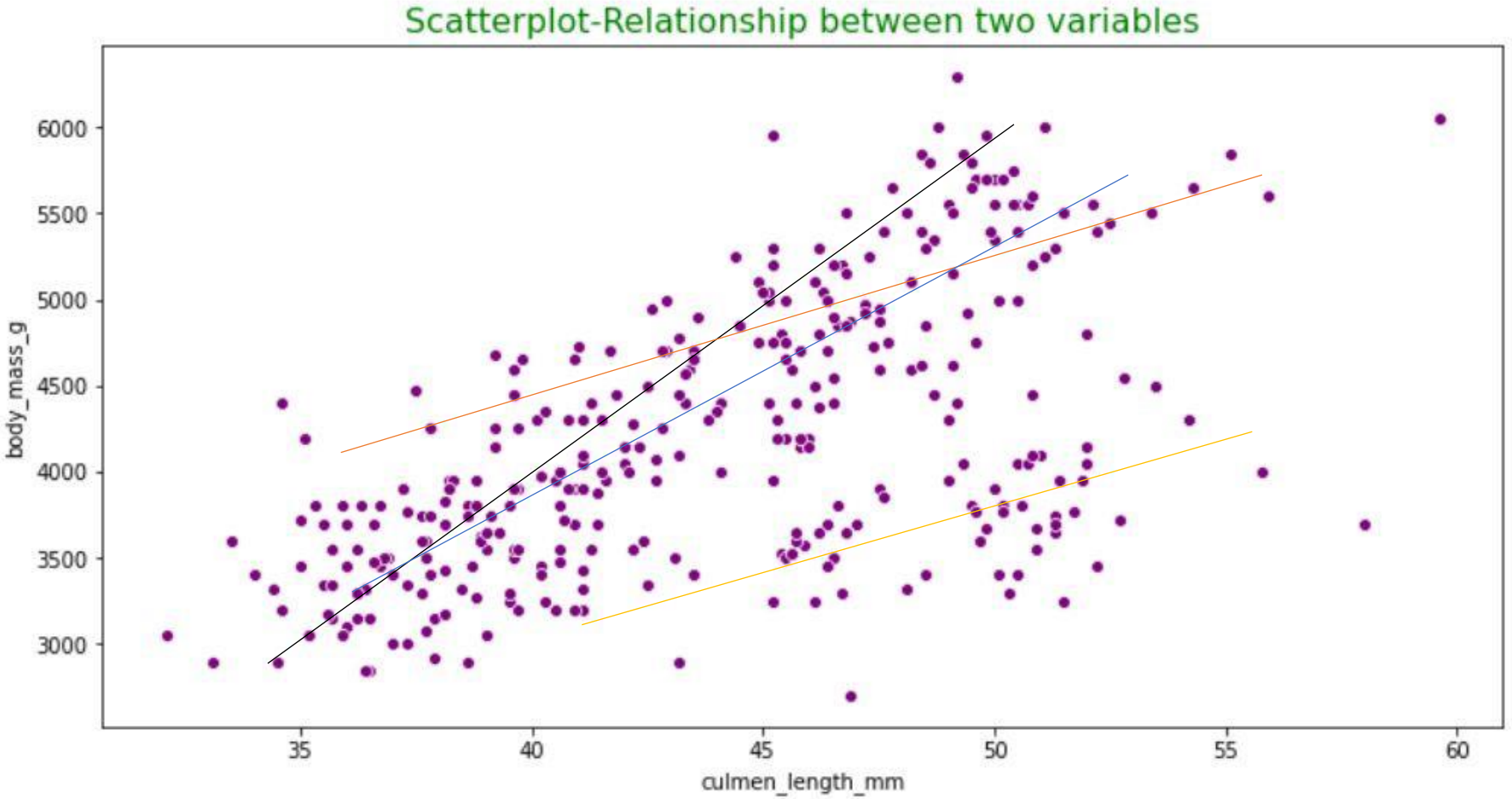
I am wondering how to choose a predictive model after doing K-fold cross-validation.

This may be awkwardly phrased, so let me explain in more detail: whenever I run K-fold cross-validation, I use K subsets of the training data, and end up with K different models.

I would like to know how to pick one of the K models, so that I can present it to someone and say "this is the best model that we can produce."

Is it OK to pick any one of the K models? Or is there some kind of best practice that is involved, such as picking the model that achieves the median test error?

INVESTIC


Scatterplot-Relationship between two variables
body_mass_g
culmen_length_mm
6000
5500
5000
4500
4000
3500
3000
35
40
45
50
55
60

INVESTIC