Train In Data

Special Cross-Validation Schemes

Cross-Validation Schemes

The Cross-Validation methods that we discussed so far assume that the data is **independent and identically distributed**.

If this is the case, a similar distribution of data is guaranteed in each fold of the Cross-Validation Scheme.

Cross-Validation Schemes

Some data may not be independent and identically distributed:

- Grouped Data (data from same subject)
- Time Series

These datasets require a tailored cross-validation scheme.

Grouped Data

Multiple observations come from the same subject

- Medical data collected from patients, with multiple samples taken from each patient.
- Voice recognition of say, digits, where the digits are pronounced by various speakers

Grouped Data

We would like to know if a model trained on a particular set of groups generalizes well to the unseen groups.

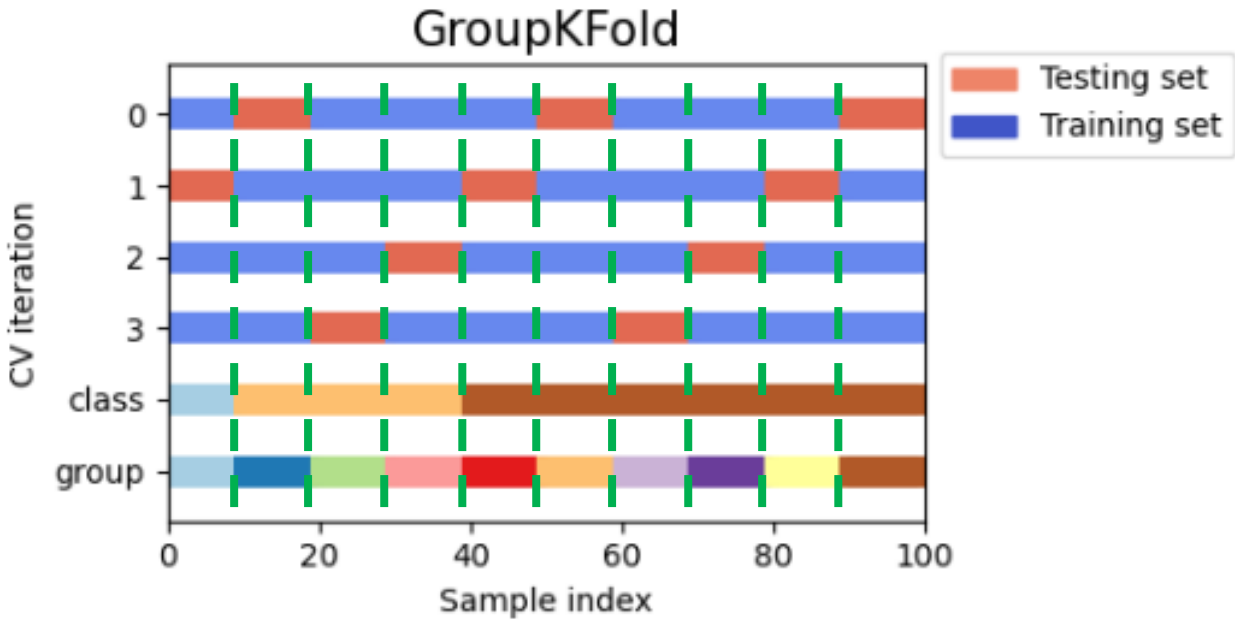
To measure this, we need to ensure that all the samples in the validation fold come from groups that are **not represented at all** in the paired training fold.

Grouped Cross-Validation

- Group K-Fold CV
- Leave One Group Out CV
- Leave P Groups Out CV

Group K-Fold Cross-Validation

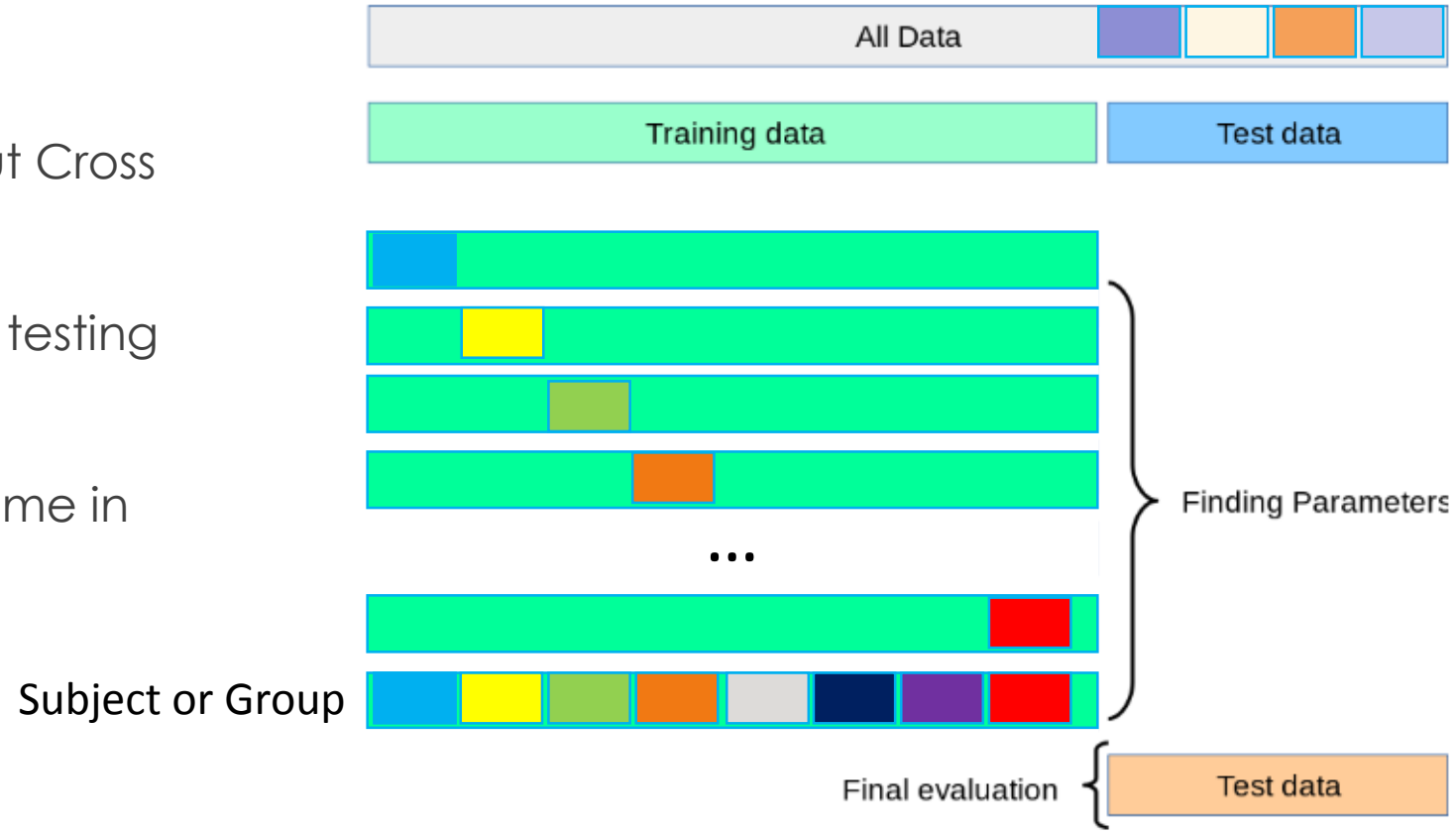
- Equivalent to K-Fold Cross Validation
- Each subject is in a different testing fold.
- Same subject is never in train and test fold at the same time
- The fold may be of different size



https://scikit-learn.org/stable/modules/cross_validation.html

Leave One Group Out CV

- Equivalent to Leave One Out Cross Validation
- Each subject is in a different testing fold.
- We leave one subject at a time in the validation fold

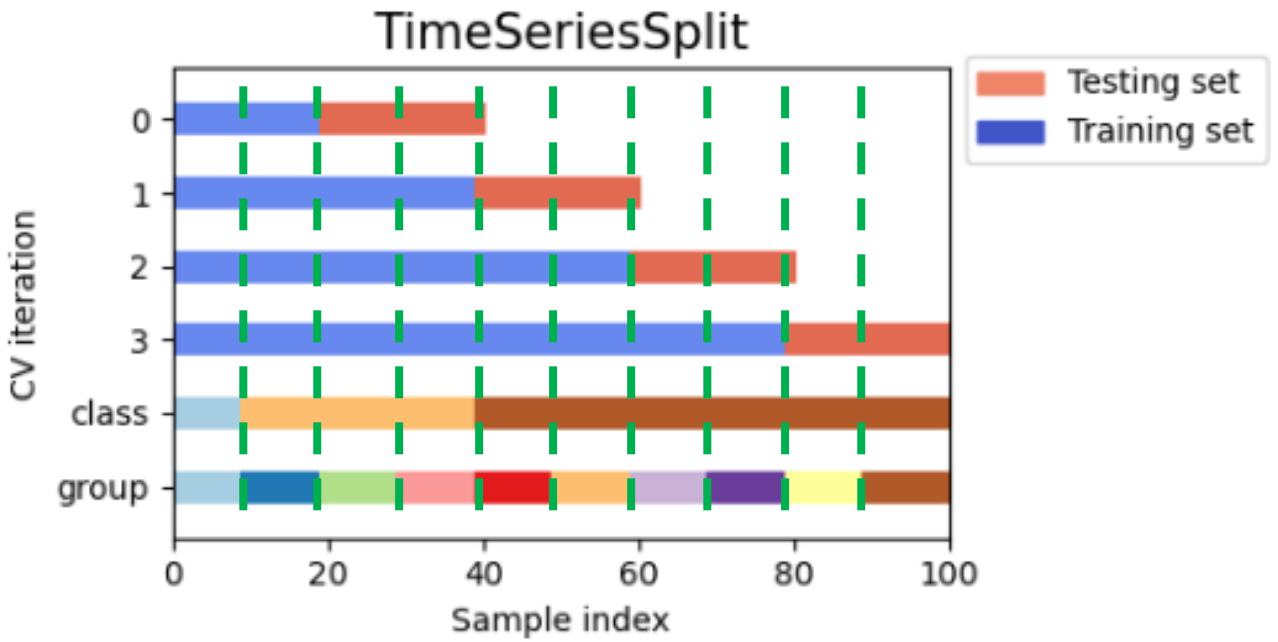


Leave P Groups Out CV

- Leaves out all possible combinations of p groups

Cross-Validation for Time Series

- We want to evaluate performance in “future” observations
- Variation of K-fold Cross-Validation
- Returns first k folds as train set and the kth fold as test set
- Unlike standard cross-validation methods, successive training sets are supersets of those that come before them.



https://scikit-learn.org/stable/modules/cross_validation.html

Train In Data

THANK YOU

www.trainindata.com