DATA SCIENCE

Leandro Ferrado, Javier Lezama, Valentina Rubiolo

ACÁMICA

9 de Mayo de 2019

Validación y testeo de modelos

- train set.

Validación y testeo de modelos

- train set.
- test set.

Validación y testeo de modelos

- train set.
- test set.
- Validation set.

Validación y testeo de modelos

- train set $\rightarrow$ 80 %.

Validación y testeo de modelos

- train set $\rightarrow 80\,\%$.
- test set $\rightarrow 20\,\%$.

Validación y testeo de modelos

- train_test_split.

Validación y testeo de modelos

- train_test_split.
- X_train, X_test, y_train, y_test = train_test_split(X,y, test_side = 0.5).

Validación y testeo de modelos

- train_test_split.
- X_train, X_test, y_train, y_test = train_test_split(X,y, test_side = 0.5).
- X_train, X_test, y_train, y_test = train_test_split(X,y, stratify = y).

Métricas

- MSE.

Cross Validation

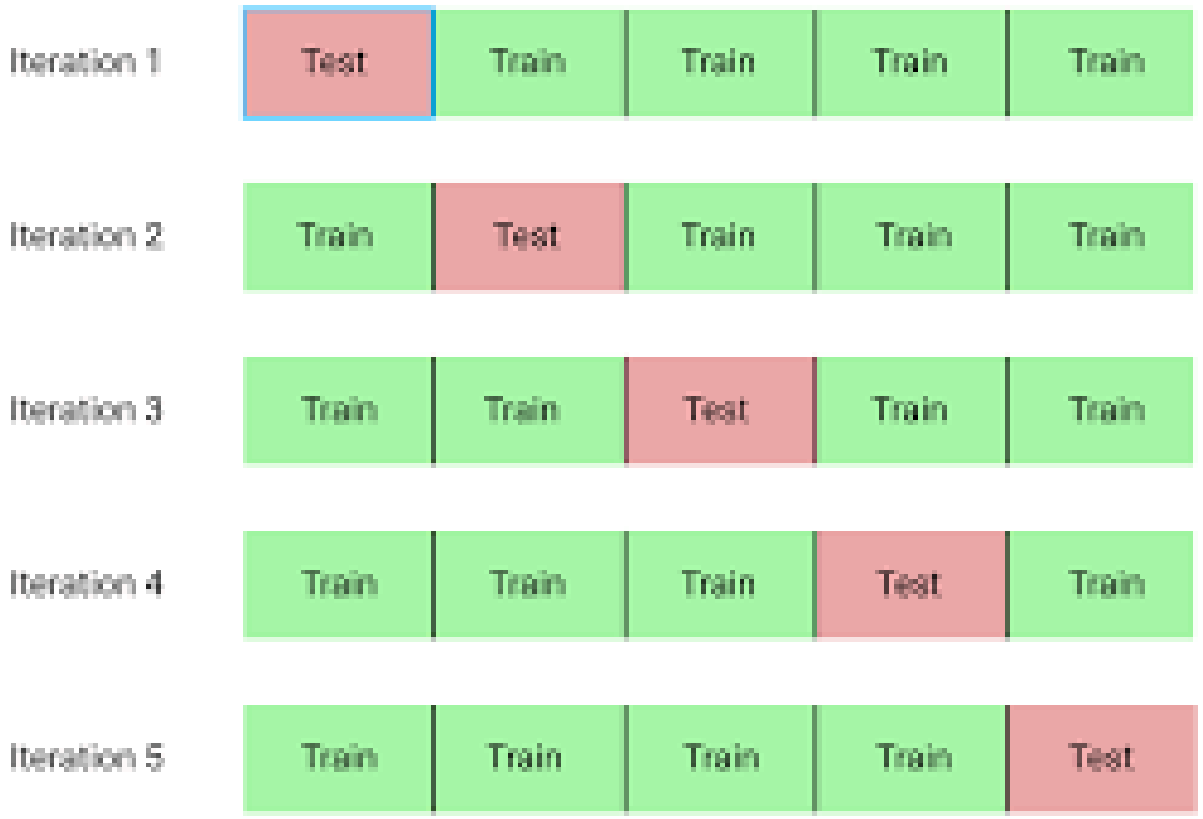
- Kfold cross validation.

Cross Validation

- Kfold cross validation.

Cross Validation

- Kfold cross validation.



Cross Validation

- random cross validation.

Cross Validation

- random cross validation.

Cross Validation

- random cross validation.

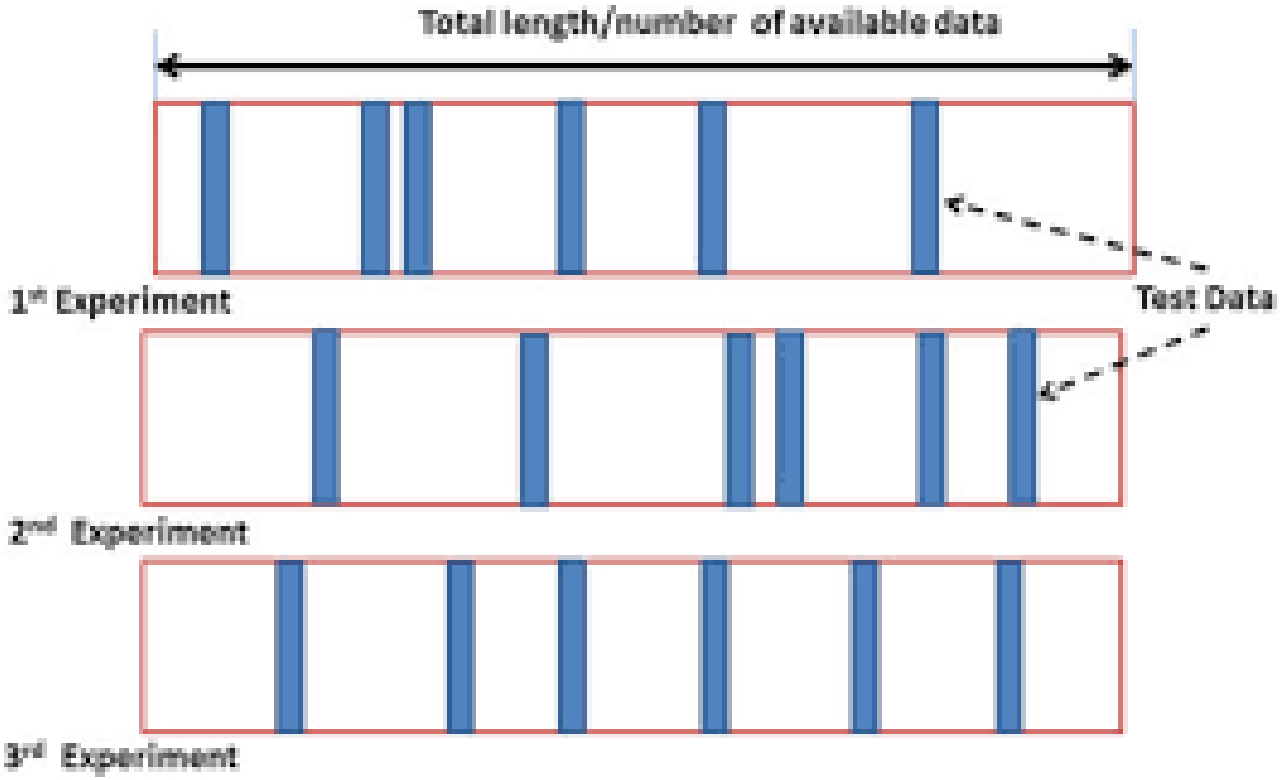


Fig. 3.7 Data splitting in the random sub-sampling approach

Cross Validation

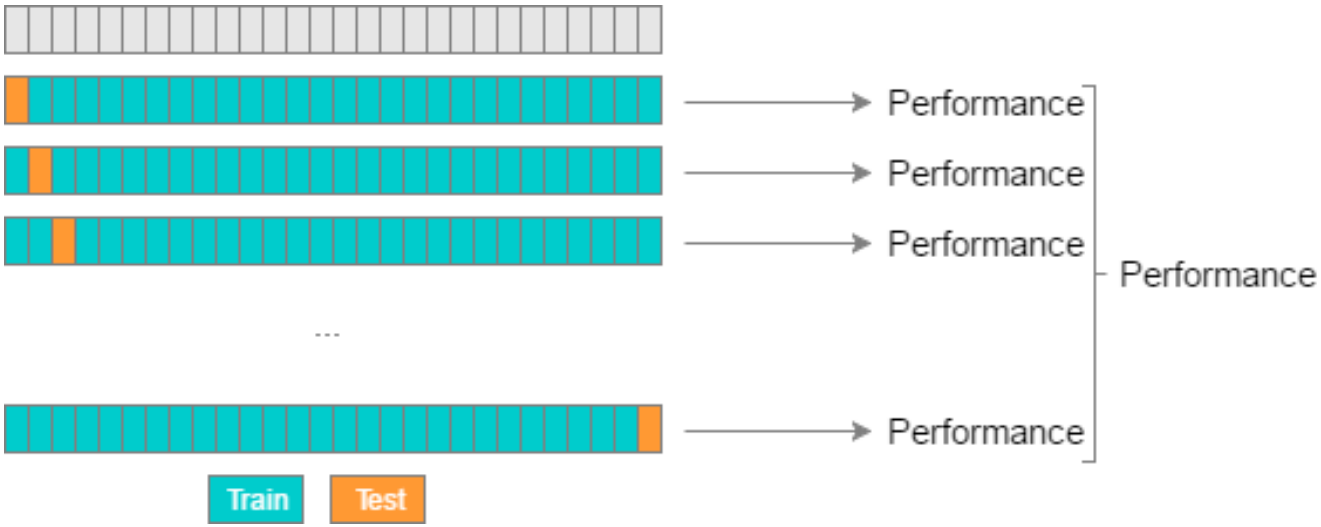
- Leave one out.

Cross Validation

- Leave one out.

Cross Validation

- Leave one out.



Cross Validation

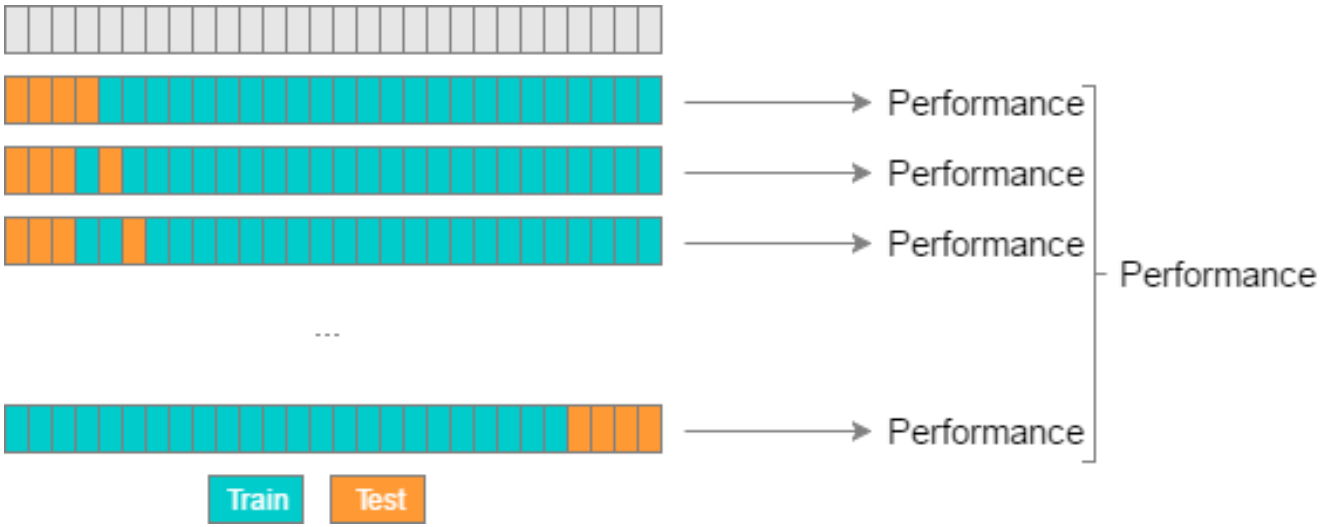
- Leave p out.

Cross Validation

- Leave p out.

Cross Validation

- Leave p out.



Cross Validation score

- crosss_val_score(modelo, X_train, y_train, scoring = métrica, cv = cantidad de iteraciones)

Métricas

- Matriz de confusión.

Métricas

- Matriz de confusión.

Métricas

- Matriz de confusión.

		Predicción	
		Positivos	**Negativos**
Observación	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Métricas

- Precisión.

Métricas

- Precisión.
- Recall.

Métricas

- Precisión.
- Recall.

Métricas

- Precisión.
- Recall.