Sane's STATS Academy of Statistics

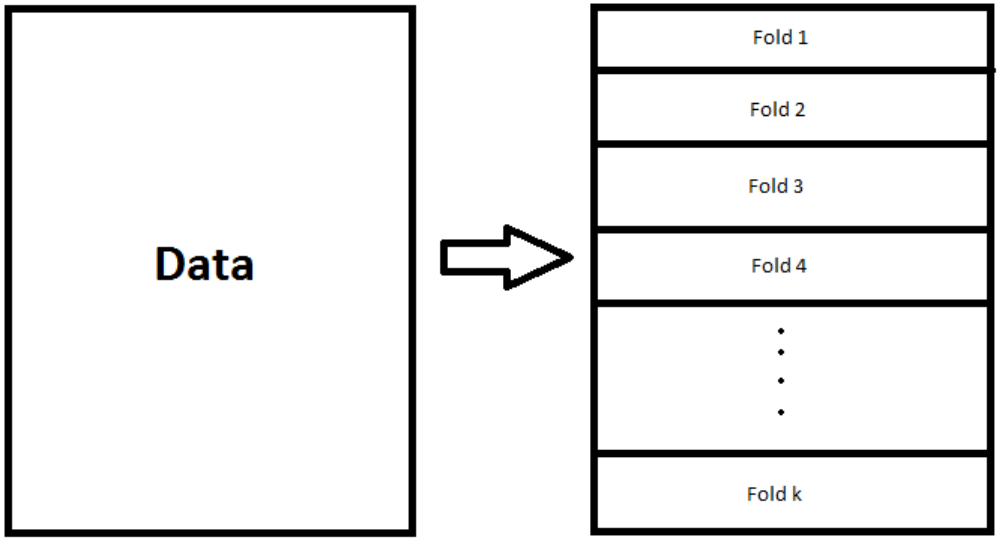
Cross-Validation

K-Folds

k-Folds

- In the cross-validation, so far we have partitioned the data into two parts or three parts
- In k-folds cross-validation, we divide the data into k equal (more or less) parts and then assuming each fold as test set average the error

K-Folds Cross-Validation



- The k-folds cross-validation process starts from dividing the data into k equal parts and that too randomly

K-Folds

	Data		Model Buliding on Training Set and Then Applying the built model on Test set				
Partitions	A		A	A	A	A	A
	B		B	B	B	B	B
	C		C	C	C	C	C
	D		D	D	D	D	D
	E		E	E	E	E	E
				Test Set			
				Training Set			

K-Folds Steps

1. Consider Fold 1 as a test set for the time being
2. With Fold 1 as test set, and remaining all other folds (data) as train set fit a model
3. Evaluate the model for the test set (Fold 1) and record the error in the evaluation
4. Consider Fold 2 as a test set
5. Repeat steps 2 and 3 for Fold 2
6. For each of the k folds, repeat 2 and 3
7. Aggregate the errors calculated in all the evaluations.