

Boostrapping

From a dataset with N observations

1. Randomly select (with replacement) N examples and use this set for training
 - ▶ remaining examples not selected for training are used for testing
2. Repeat this process for a specified number of folds K
3. The true error is estimated as the average error rate on test examples

data

 train
 test

X ₃	X ₄	X ₅	X ₃	X ₂	...	
X ₂	X ₁	X ₂	X ₅	X ₅		
X ₂	X ₁	X ₄	X ₃	X ₄		
X ₁	X ₄	X ₅	X ₃	X ₁		
X ₃	X ₁	X ₄	X ₅	X ₅		
X ₄	X ₂	X ₁	X ₁	X ₃		
X ₅	X ₃	X ₃	X ₂			
	X ₅		X ₄			

Bootstrapping

From a dataset with N observations

1. Randomly select (with replacement) N examples and use this set for training
 - ▶ remaining examples not selected for training are used for testing
2. Repeat this process for a specified number of folds K
3. The true error is estimated as the average error rate on test examples

data		<div><div></div> train</div> <div><div></div> test</div>			
X ₃	X ₄	X ₅	X ₃	X ₂	...
X ₂	X ₁	X ₂	X ₅	X ₅	
X ₂	X ₁	X ₄	X ₃	X ₄	
X ₁	X ₄	X ₅	X ₃	X ₁	
X ₃	X ₁	X ₄	X ₅	X ₅	
X ₄	X ₂	X ₁	X ₁	X ₃	
X ₅	X ₃	X ₃	X ₂		
	X ₅		X ₄		

Comparison CV and Bootstrapping