## 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

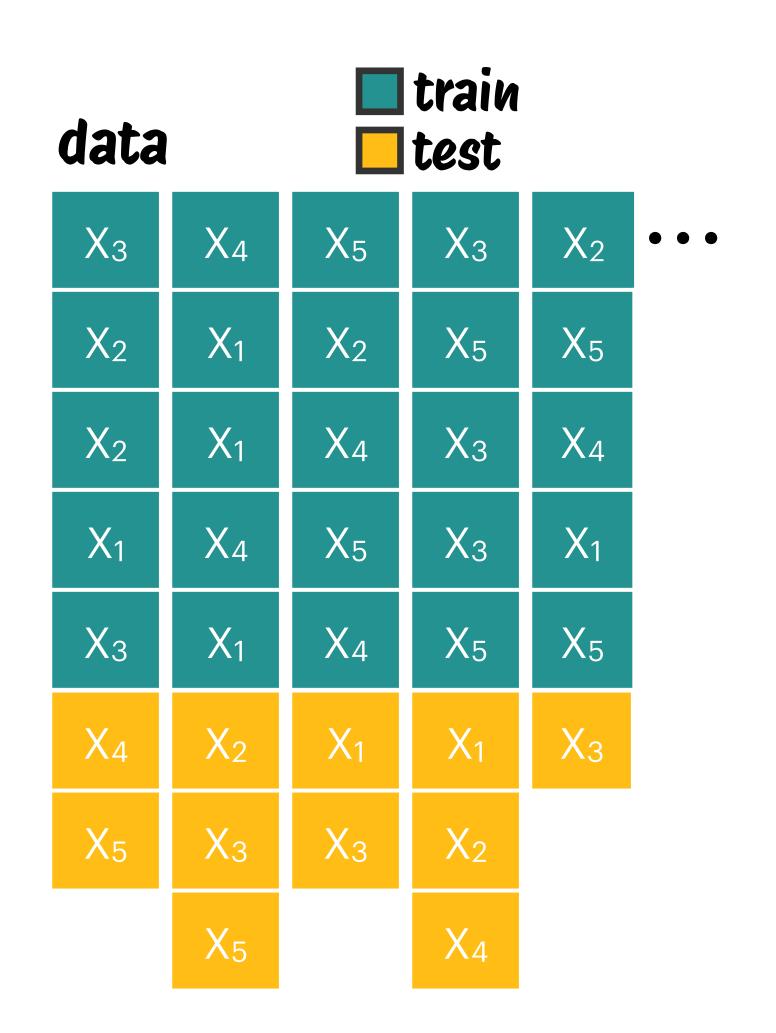
# ■ train ■ test

<b>X</b> 3	<b>X</b> <sub>4</sub>	<b>X</b> <sub>5</sub>	<b>X</b> 3	$X_2$	•
$\chi_2$	X <sub>1</sub>	$\chi_2$	<b>X</b> 5	<b>X</b> <sub>5</sub>	
$\chi_2$	X <sub>1</sub>	X <sub>4</sub>	<b>X</b> <sub>3</sub>	X <sub>4</sub>	
X <sub>1</sub>	<b>X</b> <sub>4</sub>	<b>X</b> <sub>5</sub>	<b>X</b> <sub>3</sub>	X <sub>1</sub>	
<b>X</b> <sub>3</sub>	X <sub>1</sub>	<b>X</b> <sub>4</sub>	<b>X</b> <sub>5</sub>	<b>X</b> <sub>5</sub>	
X <sub>4</sub>	$\chi_2$	X <sub>1</sub>	X <sub>1</sub>	X <sub>3</sub>	
<b>X</b> <sub>5</sub>	X <sub>3</sub>	<b>X</b> <sub>3</sub>	X <sub>2</sub>		
	<b>X</b> 5		X <sub>4</sub>		

### 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



# Comparison CV and Bootstrapping