

[← Go Back to Machine Learning](#)[:≡ Course Content](#)

Bootstrapping

What is bootstrapping?

Bootstrapping is a statistical procedure that **resamples a single dataset** to create **many simulated samples**.

How do we perform bootstrapping?

We perform bootstrapping using sampling with replacement method, where each sample is called a "Bootstrap Sample".

In sampling with replacement, we draw an item randomly from the population and put it back into the population before drawing the next item, i.e., each item has an equal chance of getting selected every time we draw an item. Hence, one item can be selected multiple times in a particular sample.

In bootstrapping, the number of records in each bootstrap sample can be smaller than or equal to the total number of records in the original data.

Let's consider an example of a bootstrap sample on dummy data: The below image shows bootstrap samples created from the original dataset. Let's say in our case, we want 3 records in each bootstrap sample.

Original Dataset					
Index	X	Y			
1	12.3	134.5			
2	4.5	56.3			
3	7.9	63.6			
4	3.55	43.3			
5	1.1	12			
6	10.4	108.1			
7	14.8	190.6			

Sampling With Replacement →

Bootstrap Sample #1		
Index	X	Y
7	14.8	190.6
5	1.1	12
7	14.8	190.6

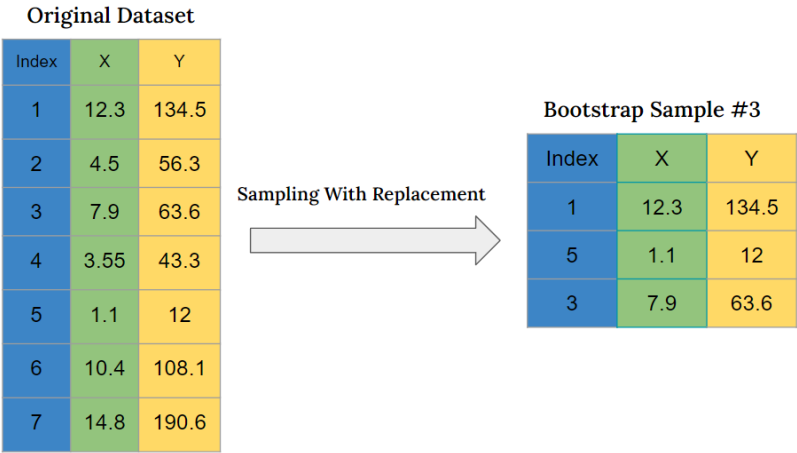
In the above image, we can see that 3 records have been randomly sampled to create a bootstrap sample and here the record with index 7 is repeated 2 times in the bootstrap sample.

Original Dataset					
Index	X	Y			
1	12.3	134.5			
2	4.5	56.3			
3	7.9	63.6			
4	3.55	43.3			
5	1.1	12			
6	10.4	108.1			
7	14.8	190.6			

Sampling With Replacement →

Bootstrap Sample #2		
Index	X	Y
5	1.1	12
5	1.1	12
2	4.5	56.3

In the above image, we can see that 3 records have been randomly sampled to create a bootstrap sample and here the record with index 5 is repeated 2 times in the bootstrap sample.



In the above image, we can see that 3 records have been randomly sampled to create a bootstrap sample and all the records are unique.