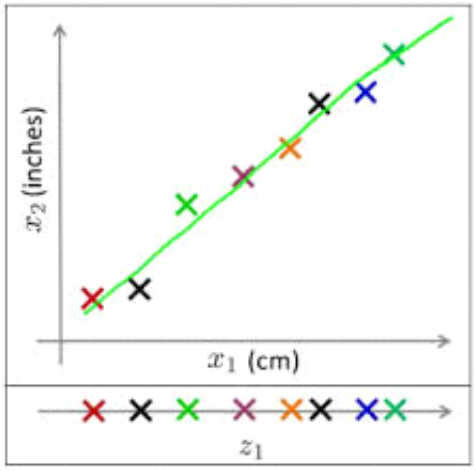
03 August 2018
18:51

What are Dimension Reduction techniques?

Dimension Reduction refers to the process of converting a set of data having vast dimensions into data with lesser dimensions ensuring that it conveys similar information concisely. These techniques are typically used while solving **machine learning problems** to obtain better features for a classification or regression task.

Let's look at the image shown below. It shows 2 dimensions x1 and x2, which are let us say measurements of several object in cm (x1) and inches (x2). Now, if you were to use both these dimensions in machine learning, they will convey similar information and introduce a lot of noise in system, so you are better of just using one dimension. Here we have converted the dimension of data from 2D (from x1 and x2) to 1D (z1), which has made the data relatively easier to explain.



In similar ways, we can reduce n dimensions of data set to k dimensions (k < n) . These k dimensions can be directly identified (filtered) or can be a combination of dimensions (weighted averages of dimensions) or new dimension(s) that represent existing multiple dimensions well.

common methods to perform Dimension Reduction

1. Missing values
2. Low Variance
3. Decision Trees
4. Random Forests
5. High correlation - Multicollinearity
6. Backward Elimination
7. Forward Selection