Dimension Reduction

WHAT IT IS: Often, a machine learning training dataset has hundreds, even thousands of features. This slows training and makes it more difficult to find an optimal solution during the process of model convergence. Dimension reduction is the transformation of data from a high-dimensional space (data with many features) into a low-dimensional space. When done well, this process retains the essential properties of the original data.

SEE IT:

WHY IT MATTERS:

GO DEEPER:

Random Projection

Random projection is used to reduce the dimensionality of a set of points which lie in Euclidean space. This algorithm projects data to a lower dimensional space using random linear projection. Random projection methods are known for their power, simplicity, and low error rates. Surprisingly, this type of projection preserves distances well. Two similar instances will remain similar after the projection while two different instances will remain quite different.

Wikipedia & Geron (p. 252)