Video

Bookmark this page

Principal Component Analysis

every other directions that
it's previously found.

transformed space, rather than in the original
feature space.

with the two dimensional
space, you'll only have a
maximum of

two vectors that can be

2:55 / 2:55 1.0x HD

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

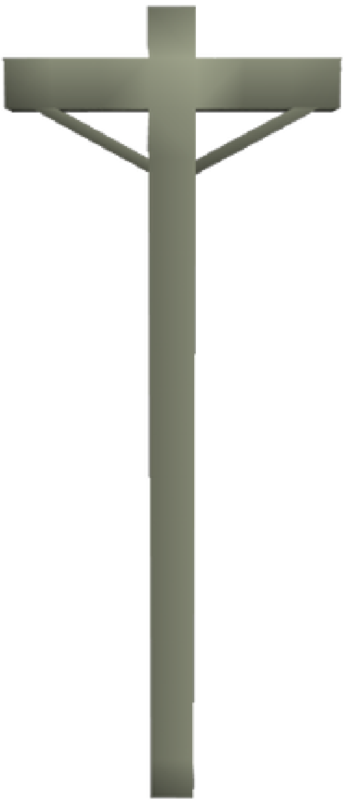
Unsupervised learning aims to discover some type of hidden structure within your data. Without a label or correct answer to test against, there is no metric for evaluating unsupervised learning algorithms. Principal Component Analysis (PCA), a transformation that attempts to convert your possibly correlated features into a set of linearly uncorrelated ones, is the first unsupervised learning algorithm you'll study.

What is principal component analysis?

PCA falls into the group of dimensionality reduction algorithms. In many real-world datasets and the problems they represent, you aren't aware of what specifically needs to be measured to succinctly address the issue driving your data collection. So instead, you simply record any feature you can derive, usually resulting in a higher dimensionality than what is truly needed. This is undesirable, but it's the only reliable way you know to insure you capture the relationship modeled in your data.

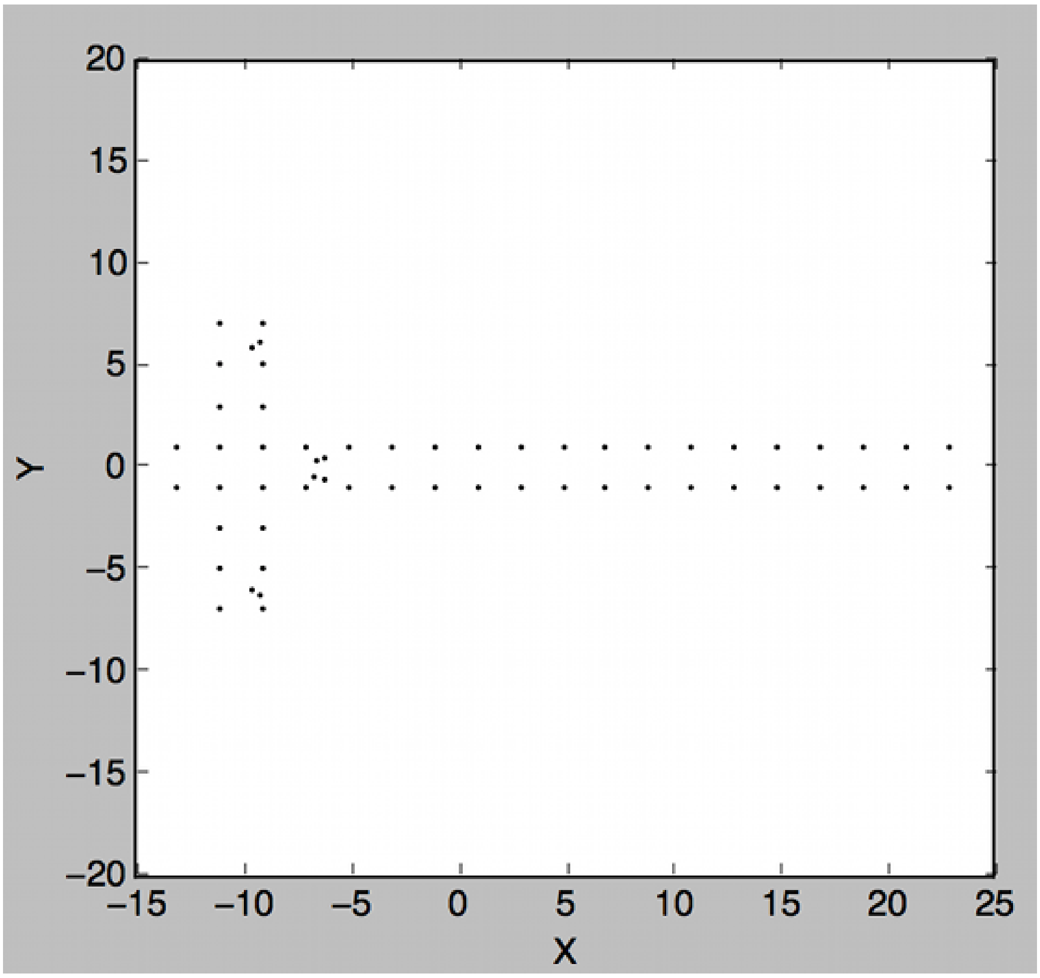
If you have reason to believe your question has a *simple* answer, or that the features you've collected are actually many indirect observations of some inherent source you either cannot or do not know how to measure, then dimensionality reduction applies to your needs.

PCA's approach to dimensionality reduction is to derive a set of degrees of freedom that can then be used to reproduce most of the variability of your data. Picture one of those cartoon style telephone poles; once you have a figure in mind, compare it to this one:



Your envisioned image probably looked similar. You could have pictured it from any other viewing angle, for instance, as if you were floating directly above it looking down:

However you probably didn't, since that view doesn't contain enough variance, or *information* to easily be discernible as a telephone pole. The frontal view, however, does. Looking at a telephone pole or any other object from various viewing angles gives you more information about that object. If the view angles are really close to one another, the information you get from the views ends up being mostly the same, with a lot of duplicate information. However if you're able to move to a completely different angle, you can get a lot more information about the object you're examining. And if you're *wise* in choose your view angles, with just a few calculated glimpses of an object, you can build a rather comprehensive understanding of it. PCA calculates those *best* view angles:



English