



Hands-On Data Science and Python Machine Learning



PREV

[Dimensionality reduction and principal component analysis](#)

Aa



NEXT

[Principal component analysis](#)

# Dimensionality reduction

So, what is the curse of dimensionality? Well, a lot of problems can be thought of having many different dimensions. So, for example, when we were doing movie recommendations, we had attributes of various movies, and every individual movie could be thought of as its own dimension in that data space.

If you have a lot of movies, that's a lot of dimensions and you can't really wrap your head around more than 3, because that's what we grew up to evolve within. You might have some sort of data that has many different features that you care about. You know, in a moment we'll look at an example of flowers that we want to classify, and that classification is based on 4 different measurements of the flowers. Those 4 different features, those 4 different measurements can represent 4 dimensions, which again, is very hard to visualize.

For this reason, dimensionality reduction techniques exist to find a way to reduce higher dimensional information into lower dimensional information. Not only can that make it easier to look at, and classify things, but it can also be useful for things like compressing data. So, by preserving the maximum amount of variance, while we reduce the number of dimensions, we're more compactly representing a dataset. A very common application of dimensionality reduction is not just for visualization, but also for compression, and for feature extraction. We'll talk about that a little bit more in a moment.

A very simple example of dimensionality reduction can be thought of as k-means clustering:



So you know, for example, we might start off with many points that represent many different dimensions in a dataset. But, ultimately, we can boil that down to K different centroids, and your distance to those centroids. That's one way of boiling data down to a lower dimensional representation.

[Support / Sign Out](#)

PREV

NEXT



Dimensionality reduction and principal component analysis	Principal component analysis
---	------------------------------