

What links tree leaves to their species?

Carson Weaver

January 22, 2022

INSERT GIF HERE: - Gif of PCA result, high-level summary

What's principal component analysis?

Principal component analysis is a way of compressing information. When you're studying something, you usually are recording lots of redundant information. An example using in THIS paper uses a ball, recorded by cameras from several different angles. All of these cameras are capturing what's basically the same information, just from different points in space. However, if you don't know what you're looking at, it's really hard to know this at the time. Principal Component Analysis, usually abbreviated PCA, is a way of extracting only the most important part of your data, usually making it smaller and much easier to draw conclusions from. In the example with the ball, you'd only get the one direction.

Ok, but what is it *actually*?

Let's dig into some of the math. PCA is described with linear algebra, meaning that it's a technique that involves vectors and matrices. Extracting the main direction that we're going to use then amounts to two steps. The first is wrapping our data up into a matrix. The next is calculating the covariance matrix from this matrix. Lastly, you extract the eigenvectors associated with the largest eigenvalues that you have. These eigenvectors are your principle components.

In our ball example, we're representing each element of our dataset as a vector.

That's cool and all, but what do eigenvectors and the covariance matrix have to do with anything?

?????

1 What kind of data even is this?

- What even is this dataset? - Motivation - Limitations - What did your first attempt look like? - Tools - Pseudocode - What's the final result? - Description of the data - Metrics describing accuracy - What did we learn about the dataset from this analysis? - How good are we at classifying leaves? - Second attempt in C++ - Tools - Translation - Performance - Visualization - Discussion - Limitations