🏠   ⌄

☰ Hands-On Data Science and Python Machine Learning

↪ Aa ☰ 🔍

Activity

As you recall from `explained_variance_ratio`, we actually captured most of the variance in a single dimension. Maybe the overall size of the flower is all that really matters in classifying it; and you can specify that with one feature. So, go ahead and modify the results if you are feeling up to it. See if you can get away with 2 dimensions, or 1 dimension instead of 2! So, go change that `n_components` to `1`, and see what kind of variance ratio you get.

What happens? Does it makes sense? Play around with it, get some familiarity with it. That is dimensionality reduction, principal component analysis, and singular value decomposition all in action. Very, very fancy terms, and you know, to be fair it is some pretty fancy math under the hood. But as you can see, it's a very powerful technique and with scikit-learn, it's not hard to apply. So, keep that in your tool chest.

And there you have it! A 4D dataset of flower information boiled down to 2 dimensions that we can both easily visualize, and also still see clear delineations between the classifications that we're interested in. So, PCA works really well in this example. Again, it's a useful tool for things like compression, or feature extraction, or facial recognition as well. So, keep that in your toolbox.