

18.3.1

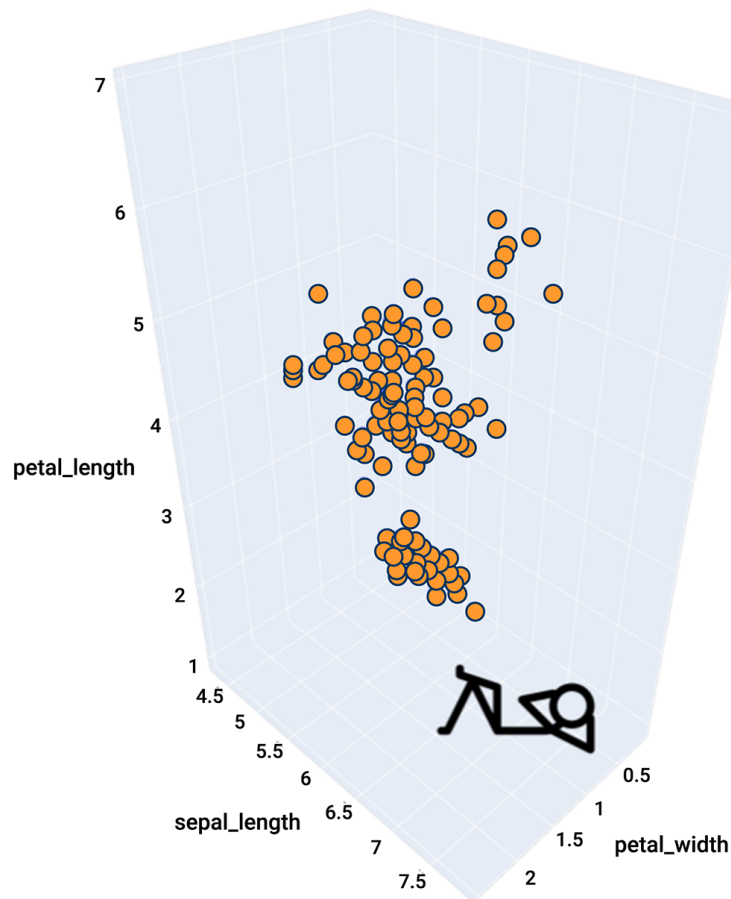
Clustering Data

Now that you and Martha are asking the right questions about your data and processing it to a state that you feel comfortable with, it's time to put it into action and start clustering!

Clustering is a type of unsupervised learning that groups data points together. This group of data points is called a **cluster**.

Imagine you are in a roomful of spheres (data points). You want to learn more about them, so you start to observe them.

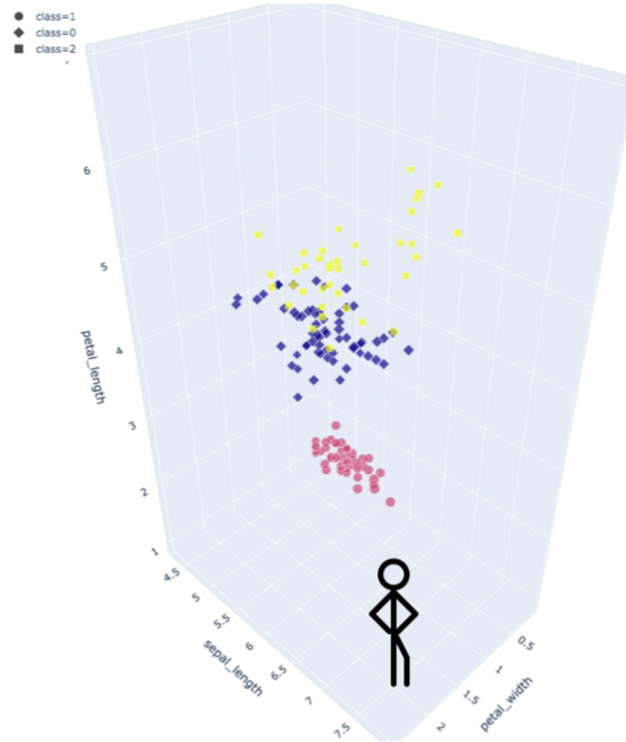
Every sphere represents a flower, and three axes represent features of flowers. After observing the flowers, you discover patterns when you combine the three features:



We can see that spheres (data points) with similar features seem to be closer together than data points with dissimilar features. We can use this spatial information to group similar data points together.

If you look at the flower features in the graph below, and start to plot them, they'll start to form groups on the graph.

After we plot the data points, they start to form three different groups, or clusters:



© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.