



Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data

▼ 5. Data Modeling

Lecture: Clustering
Quiz



Lab: Clustering
Lab



Lecture: Splitting Data
Quiz



Lecture: K-Nearest
Neighbors
Quiz



Lab: K-Nearest Neighbors

5. Data Modeling > Lab: Clustering > Assignment 4

Assignment 4

🔖 Bookmark this page

Lab Assignment 4

Feature scaling was first discussed within one of the PCA lab assignments, but this lab will really familiarize you with it. You will be making use of the **Wholesale Customer's** dataset, hosted by UCI's Machine Learning. *Unsupervised* clustering scans your features and then groups your samples based off of them. Therefore you should have a solid understanding of what each of your features are, which one's you should remove, and how to scale them in order for the 'blind' clustering to perform correctly and do what you want it to do.

1. Visit the UCI dataset page and read all the content available, so you become accustomed to the dataset. Then, load up the starter code from Module5/**assignment4.py** and as usual, read through that in its entirety as well.
2. The first thing that needs to be answered is, what is it you'd like to accomplish by clustering this dataset? There are a couple of potential questions you could ask given the data, and the one you choose will drive how you manipulate your dataset. Are you interested in which products people buy together, so that you can place them near one another in your store, or recommend the pair product when shopping online? Perhaps you're more interested in which products people are spending the most money on? Or maybe your interest just lies in identifying what individual products people are buying. All of this must be considered.
3. For the purposes of this lab, you'll assume you're interested in overall customer behavior rather than channel or region specific behavior, so you'll drop those two fields from the dataset. If you were a

Lab

**Lecture: Regression**

Quiz

**Lab: Regression**

Lab

**Dive Deeper**

▶ 6. Data Modeling II

▶ 7. Evaluating Data

▶ Course Wrap-up

large wholesaler with branches all over the nation, you'd want to keep those fields in so that you can duplicate the process outlined in the assignment to particular areas and vertical markets.

4. Complete the assignment and answer the questions below.

Lab Question

1 point possible (graded)

Which of SciKit-Learn's preprocessors causes the principal components to fan out as much as possible in an arrangement unlike the others?

☐ StanadardScaler☐ MinMaxScaler☐ Normalize☐ Scale☐ No scaling necessary**Submit**

You have used 0 of 2 attempts



© 2016 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

