



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: K-Nearest Neighbors > Assignment 5

Assignment 5

Bookmark this page

Lab Assignment 5

Remember that wheat dataset you used while exploring visualizations? It's about to make a comeback! While learning the many classification algorithms we're going to cover in the next few sections, it's a good idea to have a 'benchmark dataset' to come back to, so you can compare the performance and accuracy of other algorithms.


1. Start by looking through the starter code `/Module5/assignment5.py` and `/Module1/Datasets/wheat.data`
2. Complete the assignment except for the bonus instruction.
3. Try experimenting with other feature scaling methods, in addition to `normalize()`, to see how they affect the decision boundary.
4. Then, answer the following questions.

Lab Questions 1

1 point possible (graded)

Please enter a numeric value (e.g. 0, 1, 10.5, etc) which correctly answers the question(s) below:

What is the accuracy score of your KNeighbors Classifier when K=9 (Enter as a decimal)?

Lab 

Lecture: Regression

Quiz 

Lab: Regression

Lab 

Dive Deeper

▶ 6. Data Modeling II

▶ 7. Evaluating Data

▶ Course Wrap-up

Submit

You have used 0 of 2 attempts

Lab Questions 2

1 point possible (graded)

Decrease K by 1 and record the new accuracy score. Keep doing this until you get down to, and including, K=1. Concerning the scores you saw:

Submit

You have used 0 of 3 attempts

Congratulations on training your computer to identify wheat kernels! As you know, PCA throws away some of your data. Yet, you were able to get the high accuracy level you got in this lab by applying KNeighbors to just **two** principal components! If you're ready for a bonus experiment, remove both the PCA code as well as the visualization code from the lab. Run the KNeighbors Classifier on your entire X_train dataset and see how it performs compared to the PCA-only version you just completed above. Does it perform better? Or worse?

You can also try properly encoding the wheat_type series as a dummy feature, spanning three columns. If you attempt that, be sure to adjust your .predict() and .score() methods to fit.



© 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.

