CST 383: Intro to Data Science
Dr. Glenn Bruns

Reading: Intro to machine learning

Purpose: In this assignment you'll start to learn some of the core concepts that you'll see again and again in data science.

Instructions: Read Chapter 1 of our text 'Introduction to Machine Learning with Python'.

Answer the following questions by editing the file chap1.txt with a text editor. Be sure that you save your edited file as a .txt file -- not a pdf file, Word file, etc.

1. What do the rows in the Iris data set represent?
    a. individual Iris flowers
    b. attributes of Iris flowers
2. How many attributes of Iris flowers are found in the Iris data set?
3. What is another name for "feature"?
    a. instance
    b. sample
    c. property
4. How many Iris flowers are found in the complete Iris data set?
5. How many Iris flowers are in the training data set?
6. What is the sepal length (in centimeters) of the first flower in the Iris data set?
7. Suppose I use *accuracy* to measure how well my machine learning algorithm performs. Then what kind of machine learning am I using?
    a. supervised learning
    b. unsupervised learning
8. Using the k-nearest neighbors classifier with n_neighbors = 1, what accuracy was computed on the test data set?
9. Why does the author compute the accuracy using test data, but not the training data?
    a. The training data is much larger, so it would be harder to compute.
    b. We want to know how well the classifier will work on Irises not seen before.
10. The k-nearest neighbors classifier predicts the type of an Iris by looking at the type of Irises with similar feature values. How many such "neighbor" Irises does the classifier in the chapter use?
    a. 1
    b. 3
    c. 5

Submission. Submit your edited chap1.txt on Canvas.