Python with Applications, PIC16
E. Ryu
Spring 2019

UCLA

Homework 5
Due 5pm, Friday, June 7, 2019

Name your file `hw5.py` and submit on CCLE. Comment your code adequately.

In the 115th United State Congress, from January 3, 2017, to January 3, 2019, the Senate held 599 votes with 105 senators. Although there are at most 100 senators at any given time, 105 individuals served as senators during the 115th session; Sessions, Franken, Strange, Cochran, and McCain left the Senate before January 3, 2019, and were respectively succeeded by Strange, Smith, Jones, Hyde-Smith, and Kyl. The 599 votes are labeled by their "rollnumbers", numbered chronologically from 1 through 599. The file `senate_votes.csv` contains the voting record. A record of 1 means yes, 0 means no, and 0.5 means abstention. To learn about how to read comma-separated values (CSV) files, you can watch `https://www.youtube.com/watch?v=q5uM4VKywbA`

Problem 1: King and Sanders (entries `39` and `94` in our dataset) are independents, i.e., they are not formally affiliated with the Democratic or Republican party. For both senators, list the top 5 senators with the most similar voting records. (In this top 5 list, do not include themselves.) With which of the two major parties do their voting records best align? Use k-nearest neighbor classification with $k = 5$ to solve this problem.

Hint. You may want to use the method `kneighbors` of class `KNeighborsClassifier`.

Problem 2: Use k-means clustering, with $k = 2$, to cluster the senators into 2 groups based on their voting records. Does the result agree with the actual party affiliations? Are there any anomalies? If so, please explain.

Hint. The result of k-means clustering will vary each time you run the algorithm. To identify which of "cluster 0" or "cluster 1" corresponds to the Republican party, you can use the (Republican) Senate Majority Leader Mcconnel (entry `35` in our dataset) as a senator with a reliably Republican voting record.

Problem 3: California is represented by Harris and Feinstein (entries `11` and `12` in our dataset). Train a support vector machine on the 101 senators excluding Harris, Feinstein, and the 2 independents, and then use it to predict the party affiliation of Harris and Feinstein. Does the result agree with the actual party affiliations? Use the parameters `svm.SVC(gamma = 0.001, C= 100)`.