# ACA Assignment 4 – Music Genre Classifier in KNN

金磊

念中林

A2.

To break to tie when there exists equal distance, the algorithm we implemented will just pick first K points has minimum distance, it depends on the sorting algorithm that whether the order of same value data are kept or not.

We chose this approach for sake of convenience, since we are dealing with floating points, the probability of getting the equal distance case is pretty low. Better choice are randomly choose a winner from tied points or take all tied points into account with same weighting.

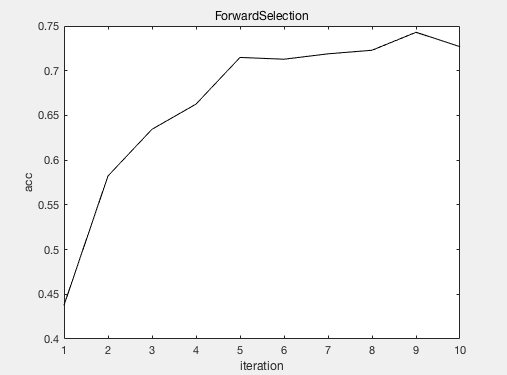
B2.

We run CV on different features, and find out that the single feature which performs the best is the 9th feature: ***Spectral Flux STD*** with accuracy around 0.45

C2.

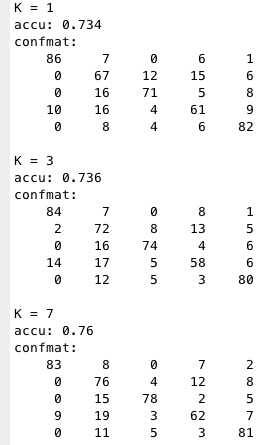
The plot of accuracy change v.s. feature selection iteration is shown below. We can tell that using 9 features can achieve relative high accuracy, but this result depends on the random seed (state) of shuffle in CV. The features will select are:

[9 10 8 3 7 1 6 2 4]



D1.

The accuracy and confusion matrix are shown blow:



1. Classical is confused with Country and Jazz
2. Country is confused with Jazz, HipHop and Metal
3. HipHop is confused with Country
4. Jazz is confused with Country and Classical
5. Metal is confused with Country

In terms of accuracy, it increases as K increases, but it's not always true as K goes much larger.

E2.

The result accuracy of K-means is about 0.866 taking the averaged 10 result , which is surprising high. This shows that music genre is closely related to music similarity.