ML report 2

TODO:

I.          K-means Problem

You will get a dataset (**data\_noah.csv**). It is Noah Syndergaard’s pitches that have been tracked by the PITCHf/x system in the MLB Regular Season.

You have to do the following:

1.            Dataset including 1321 number of instances with many attributes.

**2.**            ***Don’t*use the library related to K-means.**(i.e. Construct a K-means function by yourself).

3.            Use **Attribute x** (horizontal movement)and**y**(vertical movement) to partition 1322 pitches into 3 clusters.

4.            3 clusters will represent FF (four-seam fastball), CH (changeup) and CU (curveball).

**5.**            **Construct a cost function to check the accuracy of pitch types.**

6.            **Generate a figure** to show the result of K-Means clustering.

7.            Try to use another two or more attributes (like speed) to partition.

8.            Try to explain why k = 3 is the best, and write in your report.

9.            Show your **code**, **accuracy**, the reason of k = 3 and the result of K-Means clustering **(figure) in your report**.

II.        Kd-tree Problem

You will get a set of points (**points.txt**) in the unit square (all points have x-coordinates and y-coordinates). You have to build a 2d-tree.

You have to do the following:

**1.**            **You *can* use the library related to Kd-tree.**

**2.**            **Draw a 2d-tree divides the unit square (Use two colors).**

3.            Show your **code** and the result of 2d-tree**(figure) in your report**.