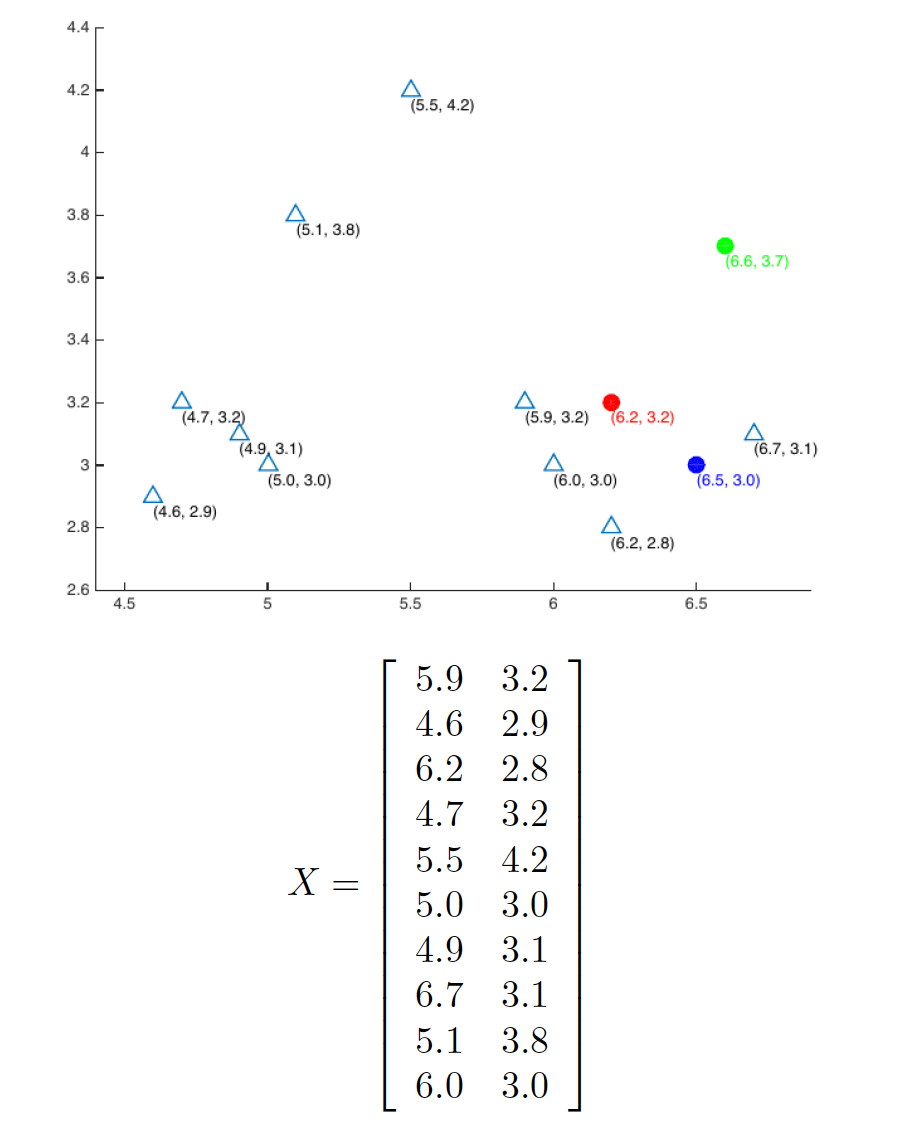
K-means Clustering



Given the matrix X whose rows represent different data points, you are asked to perform a

k-means clustering on this dataset using the Euclidean distance as the distance function. Here k is

chosen as 3. All data in X were plotted in above Figure. The centers of 3 clusters were initialized

as u1 = (6:2; 3:2) (red), u2 = (6:6; 3:7) (green), u3 = (6:5; 3:0) (blue).

Implement the k-means clustering algorithm (you are only allowed to use the basic numpy

routines to implement the algorithm).

1. Classify N = 10 samples according to nearest ui(i = 1; 2; 3). Plot the results by coloring the

empty triangles in red, blue or green. Include the classification vector and the classification

plot (task3 iter1 a.jpg).

(a) [Hint:] Using plt.scatter with edgecolor, facecolor, marker and plt.text to plot the figure.

2. Recompute ui. Plot the updated ui in solid circle in red, blue, and green respectively. Include

the updated ui values and the plot in the report (task3 iter1 b.jpg). (1pt)

3. For a second iteration, plot the classification plot and updated ui plot for the second iteration.

Include the classification vector and updated ui values and these two plots (task3 iter2 a.jpg,

task3 iter2 b.jpg) in the report. (1pt)