

HW 5 – Clustering

In this homework, you conduct clustering on the dataset.

Dataset has 8 features; the sixth variable is the categorical variable, the others are numerical variables. Please do following step and report the result (screenshot).

0. Do proper data pre-processing. You can do proper transformation for the sixth categorical variable or you can normalized the other numerical values.

1. Apply k-means algorithm to the dataset with $k=5$ and report the assignment result and the loss. You can use pre-implemented library like scipy:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.vq.kmeans.html>

2. Find the proper k (number of the clusters) with the elbow method:

[https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)#:~:text=In%20cluster%20analysis%2C%20the%20elbow,number%20of%20clusters%20to%20use.](https://en.wikipedia.org/wiki/Elbow_method_(clustering)#:~:text=In%20cluster%20analysis%2C%20the%20elbow,number%20of%20clusters%20to%20use.)

Draw the k -loss graph and find the proper k . Report the assignment result with the chosen k .