

Instructions on Weka Usage for Project 3

UE 141 on Data Mining Spring 2013

Documentation can be found from: <http://www.cs.waikato.ac.nz/ml/weka/documentation.html>
Especially you can find a manual under the Weka folder after you install Weka.

You need to use Weka Explorer. The following steps should be followed when you run K-means experiments:

1. Loading data: Open file->choose the file to be open->Open
2. Click the Cluster tab
3. Choose SimpleKmeans as the clusterer
4. Click SimpleKmeans and show the options, among which you can find numClusters (the number of clusters). The default is 2. You need to change this number to 3 or 4 in the later trials. After you change the number of clusters, you need to click OK.
5. Click start.
6. After the clustering is done, check the right panel for clustering results. Find the within cluster sum of squared error and report it.
7. At the bottom of left panel, you will find result list (right-click for options). Right click on the SimpleKmeans experiment you just run. Click Visualize cluster assignments. On the visualization interface, you can choose different X and Y coordinates to observe the clustering results. Different colors represent different clusters.

The following steps should be followed when you run hierarchical clustering experiments:

1. Loading data: Open file->choose the file to be open->Open
2. Choose the remove attribute filter:
Filter->Click Choose->Under weak-filters-unsupervised-attribute-remove
3. Click Remove. For attributeIndices, enter the indices of the attributes you want to remove. For the weather data set, if you want to keep temperature, then enter "1,3,4,5", if you want to keep humidity, enter "1,2,4,5". Click OK.
4. Click the Cluster tab
5. Choose HierarchicalClusterer as the clusterer
6. Click HierarchicalClusterer and show the options, among which you can find numClusters (the number of clusters). The default is 2. You need to change this number to 1. After you change the number of clusters, you need to click OK.

7. Click start.

8. After the clustering is done, check the bottom of left panel, you will find result list (right-click for options). Right click on the HierarchicalClusterer experiment you just run. Click Visualize tree. Now you can find the order of merging.

Datasets can be found at:

-The folder "data" under Weka folder on your machine

- <http://repository.seasr.org/Datasets/UCI/arff/>

- <http://archive.ics.uci.edu/ml/datasets.html> (You should store the data as *.csv file)