Australian Dataset

We have chosen NSW, because New South Wales is the most populated Australian state with 6 889 100, followed by Victoria with 5 205 200, and Queensland with 4 182 100.

Finding optimal number of clusters

In order to find optimal number of clusters we analyze time period of 3 years, For Australian Dataset the training period is from Jan 1, 2007 to 5, Dec 2011 and the testing period is from 6 Dec , 2011 to 5 Jan 2012.

The silhouette function is applied for the training dataset. On Fig. 1 the relationship between number of clusters and the mean value of silhouette function is represented. The peaks on the graph correspond to optimal number of clusters when the best possible clusterization is achieved, that is, distance between same cluster’s elements is minimized and distance between elements of different clusters is maximized.

Add to this figure 2 other datasets +same figure for load data

Similar figures for k-medoids, hierarchical and fuzzy-logic clusterings. (Or consider to draw only max. points in one figure)

The NY Load data is taken from Jan 1, 2007 to Dec 01, 2012. chosen



NY Price Data kmeans

