Ex5: Clustering

- Experience two major clustering "methods".
- Investigate a way to determine # of clusters,
  and evaluate its performance.

1. Homework assignment #1.

please send one pdf file, e.g., "o5_19M12345.pdf" via email to
Suzukakedai: watanabe.o.aa-cd18s@ml.m.titech.ac.jp
Ookayama: watanabe.o.aa-cd18o@ml.m.titech.ac.jp
before week6lect *of each campus*

2. Some explanation on Weka.

*1 Weka is constructed and provided by the University of Waikato
*2 For materials, see http://tcs.c.titech.ac.jp/DataMining/index.html

1. Homework assignment #5: Task

Your task:
Experience two major clustering "methods" and investigate a way of determining # of clusters.

(a) Use two data sets (bmw.arff, customers.arff), try two major learning algorithms (i.e., k-means and EM) to obtain clusters.

(b) Consider a way (possibly several ways) for
   (i) determine an appropriate # of clusters, and
   (ii) discuss its appropriateness.
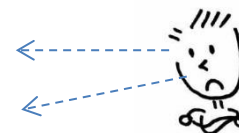
Honestly speaking,
I do not have a definite answer.

1. Homework assignment #4: Report

submit via email *before* week6lect

Required items that you need to explain:

State the results of experiments and explain your way(s) investigated at Task (b).

2. Tips for using Weka

How to add cluster values as an attribute:

・For investigating the clustering result in detail, you might want to add the cluster value (i.e., which cluster does the instance belong to) to each instance． Here is one way.

(1) check here for (2).

(2) right click here to visualize the result.

(3) save the result as an arff file.