

Homework 2

Due November 9, 2011

1. Design a Java class named `RandomPoints` that receives as input:
 - (a) a number n of points;
 - (b) the dimensionality p of the space \mathbb{R}^p in which these points reside;
 - (c) the coordinates (c_1, \dots, c_p) of the center \mathbf{c} around which these points are grouped;
 - (d) the size d of the cube centered in \mathbf{c} that contain these points.

The class should generate a $n \times p$ matrix such that each row contains the coordinates of a point. If $\mathbf{x} = (x_1, \dots, x_p)$ is one of these points then x_i should be uniformly distributed in the interval $[c_i - d/2, c_i + d/2]$.

2. For $p = 2$ use the class `RandomPoints` to produce 6 groups of points (each group containing 40 points) and draw these groups.
3. Using system *R* (which you can download for free from “Project R”), and the package `cluster` of this system, apply hierarchical clustering to the set of 240 points produced earlier using a variety of methods (single link, complete link, Ward, average distance). Also, use the k-means method for $k = 5, 6, 7$ Compare the resulting clusterings.
4. Design a clustering algorithm for grouping one-dimensional points (essentially numbers on the real line). Only the pseudocode is required and you should take advantage that the points to be clustered are unidimensional. The implementation is optional.