

STATISTICS FOR RISK MODELING

HOME WORK 4 (assigned on Tuesday, February 13, 2018)

OBJECTIVE: To apply Kohonen clustering algorithm to the 'Adult' data set (Larose & Larose, page 559)

Download the 'Adult' data set.

Apply a 3x3 Kohonen clustering algorithm to the 'Adult' data set taking:

Input: six numeric fields, seven categorical fields, and a single Boolean field (do not include income as an input);

network width = network length = 3, and default options (for the Kohonen node) otherwise.

The algorithm will lead you to 9 clusters 00, 01, 02, 10, 11, 12, 20, 21, and 22.

Construct detailed and informative cluster profiles, complete with titles. For example, the following questions may help you in developing cluster profiles: (i) which cluster is, on the average, the most educated, (ii) which cluster has the highest proportion of people who have never married? (iii) Which cluster has the highest proportion of divorced individuals? You can give the title "Divorced" to this cluster. (iv) Which cluster has the highest values of capital gains and losses? (v) Which cluster reported the least income from capital gains? (vi) Which cluster has the highest mean age? (vii) Which cluster works the highest number of mean hours per week? (viii) Which cluster is the least educated? Feel free to frame other questions in developing the cluster profiles. Numerical summaries (mainly mean values) and bar charts will be helpful.

Submit the following: (i) a small part of the data set together with a description of the input variables, (ii) your complete code, (iii) plots, charts, and tables you made for developing cluster profiles, and (iv) cluster titles and cluster profiles (two or three sentences per cluster).

References (in addition to your textbook):

<https://cran.r-project.org/web/packages/kohonen/kohonen.pdf>

[https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Clustering/Self-Organizing_Maps_\(SOM\)](https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Clustering/Self-Organizing_Maps_(SOM))

Wehrens & Buydens: Self- and Super-organizing Maps in R: The kohonen Package, Journal of Statistical Software, October 2007, Volume 21, Issue 5, pages 1-18.

STATISTICS FOR RISK MODELING

HW 3 (Assigned on Tuesday, February 13, 2018)

OBJECTIVES: (i) Exploratory Data Analysis of the CHURN DATA (see Chapter 3), and (ii) Application of the Kohonen Clustering Algorithm to the CHURN DATA (see Chapter 20).

ASSIGNMENT:

- Read Chapter 3 (including 'The R Zone'). Reproduce (using R) the charts, figures, and tables given on pages 82-88 of Chapter 3.
- Read Sections 20.5 and 20.6 of Chapter 20 (without worrying about IBM/SPSS Modeler). Reproduce (using R) the figures and tables given in 'The R Zone' on pages 557-558.
- Submit the printouts of the charts, figures, and tables reproduced by you. There is no need to submit your R code unless your code is different from 'The R Zone' in your book. Also, there is no need to submit any discussion/interpretation of the charts, figures, and tables reproduced by you.