

CSI5387: Winter 2014

Assignment 1

Due date: February 5, 2014

(Please submit a printed copy to the instructor in class)

In this assignment, you will be exploring the effect of multi-class versus binary-class learning on two classifiers naturally able to handle these two types of data: Decision trees and Multi-Layer Perceptrons. You can use the WEKA implementations of these classifiers (you can download WEKA from <http://www.cs.waikato.ac.nz/ml/weka/>), or any other implementations such as those you can find in Matlab or R. You will be exploring the performance of these two classifiers on 1) binary problems; 2) multi-class problems whose number of classes remains moderate; and 3) multi-class problems with large numbers of classes. (Define, on your own, what you consider moderate and large and explain where and why you make that distinction. If you find that distinction in the literature, make sure you cite your sources).

Here are your tasks:

- Start by searching the literature to find out what has previously been said about the problem [You will have to cite all your sources, obviously].
- From your small literature survey, extract a hypothesis that you will set out to test. Clearly enunciate what that hypothesis is. [Note: your survey doesn't need to be exhaustive. Also, your hypothesis does not have to be new. It is appropriate, for this assignment, to verify whether a hypothesis stated by another researcher can be supported by your experiments or not].
- The next step will be to test that hypothesis.
 - Visit the UCI Repository Website (<http://archive.ics.uci.edu/ml/>) to select and download problems that you believe are appropriate to test your hypothesis. You can also generate your own artificial data sets if that helps you make your point. This is optional, however.
 - Explain what your testing methodology will be, what kind of results would constitute support for your hypothesis and what kind would constitute its refutation. Also state your expectations.
 - Run your experiments. (Watch out: Remember that Neural Nets take a long time to train and are very sensitive to the parameters you use, particularly, the number of hidden units. Make sure you experiment sufficiently with different parameters).
 - Report your results using appropriate evaluation metrics and statistical tests. You can use the R code from the Japkowicz and Shah textbook (downloadable from http://www.site.uottawa.ca/~nat/Courses/csi5387_Winter2012/BookWebsite.zip) Select the result reporting format that you judge to be the most appropriate to make

the point you wish to make (i.e., line graphs, bar graphs, tables). Make sure not to overwhelm the reader with your results and yet make sure that you are reporting enough results to make your point.

- Discuss your results. Were you able to support the hypothesis that you formulated or not? [I will be happy in either case: an assignment that manages to support it will not necessarily get a higher mark]. Why, do you believe that you obtained the results you did? i.e., what are the limitations of your study? [Every study has limitations! You must list those of yours clearly]. What can you truly say, based on your results, about the hypothesis you formulated?
- Additional observations: In the course of your study, did you encounter interesting issues that, if time permitted, you would like to explore? Please state them and explain why they are of interest to you. [Note that such issues are usually discussed in a section of your report called “Future Work”]. BTW, if you are truly interested in exploring the issues that you have uncovered further, your course project could be the place to do it!

Warnings/Notes:

- More is not necessarily Better! This assignment can take the size you want. It can be a small study or it can lead to a Ph.D. Thesis. I am not looking for a Ph.D. thesis here! Assess your time properly, bound the project accordingly and don't kill yourselves: I will not necessarily reward people who run more experiments more than people who run fewer.
- Here are the objectives of this assignment:
 - Practical experience with Decision Trees and Neural Nets
 - Practical experience with evaluation techniques
 - Practical experience searching the literature on a particular topic
 - Practical experience in formulating and testing a hypothesis and making sure that all your statements and arguments are logical.
 - Practical experience in reporting results.
- Your conclusions do not have to be ground-breaking. I am looking for some novelty in your project, but this assignment does not need to lead to new insights. (If you discover new insights, that's fine too and make sure to report them, but that is not the purpose of the assignment).
- Find a way to have fun while doing this assignment! Research is fun as long as you formulate the questions you research in the way that interests you.
- If you are not happy with this assignment (i.e., you find that it is not reasonably sized), let's talk about it in class... I like to encourage my grad students to explore ideas deeply, but I am also open to discussions, especially when I go overboard! It is possible to reach a compromise if necessary!