

10-701/15-781 Machine Learning, Fall 2003

Homework 6/Mini-project 2

Out: November 11, 2003

Due: start of class November 25, 2003

If you have questions, please contact Ning Hu <ninghu+781@cs.cmu.edu>.

You have tried formulating and solving a machine learning problem using a real-world data set in Mini-project 1. This time, you will have the opportunity to try tackling a new one. You may work with any data set as you wish. You *are* also allowed to follow up on your original project work, if you have ideas for substantial new improvements or investigations.

What you will do.

1. Find a partner to work on this project (two person teams are encouraged, though you may work alone if you prefer. No three person teams please.)
2. Choose your data set. There are several choices:
 - You may use the data sets provided in Mini-project 1, as long as it is not the one already used in the your previous mini project. But if you choose to follow up on your last project work, you may still use the original data set.
 - There are some good sources of interesting data sets:
 - UCI Knowledge Discovery in Databases Archive
<http://kdd.ics.uci.edu/>
 - UCI Machine Learning Repository
<http://www.ics.uci.edu/~mlearn/MLRepository.html>
 - StatLib—Datasets Archive
<http://lib.stat.cmu.edu/datasets/>
 - You are allowed to generate or obtain your own datasets.
3. Define your learning task according to the data set.
4. Perform the work. As a guideline, we expect each student to spend 7–12 hours on this homework over the course of two weeks (remember you are working in pairs, so you can do a fairly substantial project).
5. Turn in a short write-up. Your write-up should describe *precisely* your data source, your learning task(s), your learning method(s) including how you represented the

data for input to the learner, your experiments, results, and any conclusions you draw from this. If you choose to follow up on your original project work, you must state "Follow up" work and include a copy of your original Mini-project 1 submission. Be clear in your write-up about what you have done that is new.

The clarity and content of your write-up will have a primary impact on your grade. The reports must not be more than 5 pages, 11-point font, including figures. Each two-person team must hand in a single write-up.

Grading and determining when you have done enough.

We strongly advise that you develop a baseline learning system by the end of week 1, then work on some interesting extension or alternative during week 2. A project that does a solid job applying the given code and carefully evaluating and describing it might get 75–80% credit. A project that in addition pursues an interesting second approach or alternative problem might get 90–100% credit.

Be creative! Exploring your own interesting ideas and comparing them with the baseline approaches will receive credit whether they beat the baseline or not.

If you choose to follow up on your original project work, the grading will depend on how you design and implement your new idea, how much improvements it achieves comparing to your original baseline approach, and what conclusion you can draw from the experiments.