

CSE 291: Statistical Learning

Extra Credit Assignment

This assignment is due before 10am on Thursday March 18, 2004. Because of the final exam and the deadline for submitting letter grades, assignments must not be submitted late.

You may do any or all of the problems below. Extra credit will replace your lowest scores on previous assignments.

(1) In a concise mathematical narrative, fill in all missing steps in Section 8.5.2 of *The Elements of Statistical Learning*.

See Exercise 7.31 in *Statistical Inference* by Casella and Berger for guidance. Be sure to explain clearly the meaning of all expectations, and of each conditional or unconditional likelihood function that is involved. Choose the notation that you use carefully to maximize clarity.

(2) Continue your mathematical narrative to derive Algorithm 8.1 of *The Elements of Statistical Learning*.

(3) Formulate the AP&M elevator problem in a precise way as an application of expectation-maximization.

(4) Investigate numerically the bootstrap method for (a) generating a confidence interval for the median, and (b) testing the null hypothesis that two means are the same. Consider a range of sample sizes, and consider at least three non-Gaussian distributions: one with heavy tails, one that is skewed, and one that is multimodal.

Use your results from part (a) to investigate the efficiency of the median as an estimator of the central tendency of a symmetric distribution. For part (b), investigate the correctness of the significance level of the bootstrap test, and also its power. In all cases, try to reach conclusions that are clear, definite, general, and well-supported.

(5) Look for relevant house price data on the web, and use it to find a family of distributions that fits this type of data well. Then investigate confidence intervals for medians of distributions in this family.

Use your investigation to derive rules of thumb for evaluating the significance of median differences in house prices, as reported at <http://realestate.signonsandiego.com/mediancharts/0104sdmedian.html> by the San Diego Union Tribune for example.

See http://sd.znet.com/~schester/fallbrook/home_prices/ for a good informal discussion of median house prices, and for some examples of house price distributions.