STA 414/2104, Spring 2007 — Assignment #1

Due on February 20 by 5pm (hand in to my office, SS6016A). Or you can hand it in in class on February 16, if you're done by then. Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

In this assignment, you will implement linear basis function models, fit using maximum penalized likelihood (ie, penalized least squares), with the magnitude of the penalty chosen by leave-one-out cross-validation, and assess how well this method performs on an artificial data set with two inputs.

The data is on the course web page, `http://www.utstat.utoronto.ca/~radford/sta414`, in four files. The file `a1-x-train` contains the inputs for 50 training cases (two inputs on each of 50 lines). The file `a1-t-train` contains the corresponding target values (one target per line). The files `a1-x-test` and `a1-t-test` contain corresponding values for 361 test cases (arranged on a grid). You should fit your model entirely on the training data (including selecting a suitable value for the magnitude of the penalty, λ). Only at the end should you look at the test data, making predictions for each test case based on the inputs in `a1-x-test`, and then comparing your prediction with the actual value in `a1-t-test`.

You should use a model with 122 basis functions, each a function of the two input values for a case. The first basis function, $\phi_0(x)$, should be the constant value 1. The remaining 121 basis functions should have the form

$$\phi_k(x) \;=\; \exp\Big(-\frac{1}{2\rho^2}\sum_{j=1}^{2}(x_j - c_{kj})^2\Big)$$

where c_k is a vector giving the "centre" of basis function k, and ρ is a constant that we will set to 0.15. The centres of the basis functions should form a grid in which the two coordinates take on all values in the set $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

The penalty that you should add to minus the log likelihood should be as follows:

$$\lambda \sum_{k=1}^{121} w_k^2$$

Note that w_0, the coefficient for $\phi_0(x)$ is not included in the penalty term.

For a given value of λ, you should find the estimated value for the vector of weights, w, using equation (3.28) from the text, modified to account for the absense of w_0 in the penalty. To decide on a value to use for λ, you should consider a set of values for λ and evaluate how good each value seems to be using leave-one-out cross-validation. You'll need to play around a bit to see what a suitable set of values to consider might be. To do leave-one-out cross-validation, you will need to find 50 estimates for w, leaving out each training case in turn, and see how well each estimate predicts the left-out case. You should use average squared error for these predictions to judge which value of λ is best.

Once you have decided on a value for λ, you should make predictions for the 361 test cases, and see what the average squared error is for these predictions. You should also make predictions using some other values for λ, in order to see whether or not leave-one-out cross-validation has found close to the best value of λ. Also, for the chosen value of λ, and one that is quite a bit smaller, and one that is quite a bit bigger, you should produce contour plots of the predicted values for the test cases, in order to see how the choice of λ affects the smoothness of the predictions.

Your program should be written in R, but should not use any of the built-in R functions for fitting linear models. (Ie, you should implement the computations yourself, using R's functions for matrix operations.) You should hand in a listing of your program, with suitable but not excessive comments. You should also hand in the output of your program, and a brief discussion of the results.