

## Stat 648 Assignments

1. Problems 2.1, 2.2, 2.4, 2.6, 2.7 of HTF1 (the first edition of HTF). This is 2.1, 2.2, 2.4, 2.7 and 2.8 of HTF2
2. To get some feel for what various kinds of potentially new-to-you assumptions about functional forms for prediction methods, contour plot the following functions as well as the two summands that make each of them up.
  - i)  $g(x_1, x_2) = \exp\left(-\frac{1}{6}\left((x_1-0)^2 + (x_2-0)^2\right)\right) - 2\exp\left(-\frac{1}{2}\left((x_1-1)^2 + (x_2-1)^2\right)\right)$  (a linear combination of 2 "radial basis functions" that are bivariate normal pdfs)
  - ii)  $g(x_1, x_2) = \frac{1}{1 + \exp(x_1 + x_2)} + \frac{2}{1 + \exp(1 + 10x_1)}$  (a very simple "neural net")
  - iii)  $g(x_1, x_2) = (x_1 + x_2) + \sin x_1$  (a possible form for a model fit by "projection pursuit")
3. The regression computing assignment that Melissa sent out.
4. Problems 3.5, 3.12, 3.16, 3.30 and 4.6 of HTF2.
5. A classification computing assignment constructed by Melissa.
6. Problems 5.2, 5.7, 5.11, 5.12, 5.15 of HTF2.
7. Problems 6.2, 7.3, 7.6 of HTF2.
8. Consider the regularized fitting problem for fitting real-valued functions on  $[0,1]$ , with kernel  $K(z, x) = (1 + zx)^2 = 1 + 2zx + z^2x^2$ . Notice that for every fixed  $z \in [0,1]$ , the function  $K(\cdot, z)$  is quadratic. (This is about the simplest case of the RKHS stuff that one could make up. So functions  $\sum_{i=1}^N b_i K(\cdot, x_i)$  are polynomials of degree at most 2.)
  - i) Argue that every eigenfunction  $\phi$  of  $K$  in  $L_2[0,1]$  is quadratic. Argue that every non-zero eigenvalue  $\gamma$  of the kernel must satisfy

$$\det \begin{pmatrix} (1-\gamma) & \frac{1}{2} & \frac{1}{3} \\ 1 & \left(\frac{2}{3}-\gamma\right) & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{4} & \left(\frac{1}{5}-\gamma\right) \end{pmatrix} = 0$$

and solve this for the three positive eigenvalues of  $K$ . Use these to identify three non-zero eigenfunctions and verify that  $K(z, x) = \sum_{i=1}^3 \gamma_i \phi_i(z) \phi_i(x)$ .

ii) I'll send you a training data set of size  $N = 16$  with  $x_i = i/16$  and  $y_i = 2 \exp(x_i) + \varepsilon_i$  for  $\varepsilon_i \sim \text{iid } N(0, (.5)^2)$ . Fit to the data an element of  $\mathcal{H}_K$  minimizing

$$\sum_{i=1}^{16} (y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}_K}^2$$

for  $\lambda = .0001, .01$ , and  $1$ . What "degrees of freedom" are associated with these fits?

iii) For the cases in ii) compute the positive eigenvalues and corresponding (normalized) eigenvectors  $\mathbf{e}$  for the  $16 \times 16$  matrix  $\mathbf{K}$ . Then plot  $e_i$  vs  $x_i$  for these eigenvectors and compare to plots of corresponding  $\phi(x_i)$  vs  $x_i$ . How do eigenvectors and eigenvalues for the matrix  $\mathbf{K}$  compare to eigenfunctions and eigenvalues for the kernel  $K$ ? Redo this comparison if only data with  $i = 1, 2, 5, 6, 11, 12, 14, 16$  are available.

iv) Let  $g(x) = a + bx + cx^2 + dx^3$  be a non-zero cubic function on  $[0, 1]$  that has

$$\int_0^1 g(x) \phi_i(x) dx = 0$$

for the three eigenfunctions  $\phi_i$  in i). Argue that  $g$  is in the null space of the kernel  $K$ . Then find a constant  $\alpha$  and function  $f \in \mathcal{H}_K$  minimizing

$$\sum_{i=1}^{16} (y_i - (\alpha g(x_i) + f(x_i)))^2 + \lambda \|f\|_{\mathcal{H}_K}^2$$

for each of  $\lambda = .01, 1$ , and  $100$ .

9. I'll send you an  $(x_i, y_i)$   $N = 50$  with  $x_i = i/50$ .

i) Fit the following to the data and plot  $\hat{y}_i$  vs  $x_i$  (along with  $y_i$  vs  $x_i$ ).

a) cubic smoothing splines with associated degrees of freedom 3 and 10, and

b) locally weighted linear regression smoothers using the standard normal density for a kernel and degrees of freedom 3 and 10.

ii) Use leave-one-out cross-validation to identify a "best" choice of cubic smoothing spline for fitting these data. Then use 10-fold cross-validation to do the same, based on

the sets  $\mathcal{J}_k = \left\{ \frac{k}{50}, \frac{k+10}{50}, \frac{k+20}{50}, \frac{k+30}{50}, \frac{k+40}{50} \right\}$  for  $k = 1, 2, \dots, 10$ .

10. For the data set in Problem 9, fit and plot solutions to the regularized fitting problem of minimizing

$$\sum_{i=1}^{50} (y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}_K}^2$$

- i) for the kernel  $K(x, z) = \exp(-(x - z)^2)$  and  $\lambda = .0001, .01$ , and  $1$ ,
- ii) for the kernel  $K(x, z) = \exp(-2(x - z)^2)$  and  $\lambda = .0001, .01$ , and  $1$ , and
- iii) for the kernel  $K(x, z) = \exp(-2|x - z|)$  and  $\lambda = .0001, .01$ , and  $1$ .

What degrees of freedom are associated with each of the fits in i) through iii)?

11. Problem 9.6 of HTF2. Also fit a projection pursuit regression and a single layer feed forward neural net to these data. (Presumably the `ppr()` function in the R package `stats` and the R package `nnet` can be used to do the latter. Also presumably, the `mgcv` package, the `rpart` package, `mars()` in the `mda` package, and the `PRIM` package may be useful.) In none of this do you need to do the work of formal cross validation, but in some rough sense try to fit things with enough complexity to be comparable to the HTF analysis. Make whatever plots make sense and for which you have the time and background to cook up to illustrate the nature of your fits.

12. Izenman Problem 14.4. Consider the following 10 two-dimensional points, the first five points,  $(1, 4), (3.5, 6.5), (4.5, 7.5), (6, 6), (1.5, 1.5)$  belong to Class 1 and the second five points  $(8, 6.5), (3, 4.5), (4.5, 4), (8, 1.5), (2.5, 0)$ , belong to Class 2. Plot these points on a scatterplot using different symbols or colors to distinguish the two classes. Carry through *by hand* the AdaBoost algorithm on these points, showing the weights at each step of the process. (Presumably Izenman's intention is that you use simple 2-node trees as your base classifiers.) Determine the final classifier and calculate its misclassification rate.

13. (Do this one absolutely on your own. No discussion with others is allowed here.) Consider a neural net with 2 hidden layers. Suppose that 3 input variables pass through 2 nodes at the first hidden layer, that outputs of these hidden nodes pass through 2 nodes at the second hidden layer, and that there is a single output variable. Carefully lay out the version of the back-propagation algorithm appropriate to fitting this network.