HW3 (ECE7252 Spring 2008)

1. HTF Exercise 2.6. [Note: ignore the k-nearest neighbor part in (a), and x_0 is a new point to be considered.]
2. HTF Exercise 3.7.
3. HTF Exercise 3.10.
4. HTF Exercise 3.12.
5. An exercise on linear prediction analysis:

Consider the difference equation: $h(n) = \sum_{k=1}^{p} \alpha_k h(n-k) + G\delta(n)$

The autocorrelation function of $h(n)$ is defined as: $\tilde{R}(m) = \sum_{n=0}^{\infty} h(n)h(n+m)$

a) Show that $\tilde{R}(-m) = \tilde{R}(m)$;

b) By substituting the difference equation into the expression for $\tilde{R}(-m)$, show that

$$\tilde{R}(m) = \sum_{k=1}^{p} \alpha_k \tilde{R}(|m-k|),\ m = 1, ..., p.$$

6. Consider the first order linear prediction $\tilde{x}(n) = \alpha x(n-1)$, with *x(n)* a stationary, zero-mean signal with phi_x(1) denoting the first-order autocorrelation function. Define the prediction error as $d(n) = x(n) - \tilde{x}(n)$.

a) Show that $d(n)$ has a variance: $\sigma_d^2 = \sigma_x^2 \bullet [1 + \alpha^2 - 2\alpha\phi_x(1)/\sigma_x^2]$;

b) Show that σ_d^2 is minimized with $\hat{\alpha} = \phi_x(1)/\sigma_x^2$;

c) Show that the minimum prediction error variance is $\sigma_d^2 = \sigma_x^2 \bullet (1 - \hat{\alpha}^2)$;

d) The model here can be apply to automatic gain control in signal acquisition systems that adjust the amplification factors based on the signal levels. Can you design such an AGC mechanism?

7. Based on the prostate cancer data, regenerate some results shown in Table 3.3, and Figure 3.6 using four of the regression techniques we discussed in Chapter 3, including LS, subset selection (pick a few intelligent choices), ridge (try a few lambda values), and PCR regressions (with *M*<8). [Note: no Lasso or PLS are needed, for subset selection you can experiment with a few and do some analysis to see which subset might give you the best results.] Conduct at least the following analysis:

1) Plot some scatter diagrams of *Y* against *Xs*, and what can you tell?
2) Compute correlation coefficients between *Y* and all *Xs*, and what can you tell?
3) Compare the 4 regression models. Are there major differences? Are the residual curves in Figure 3.6 good ways to visualize the comparisons?
4) If we move two of the variables among the 67 cases to large values (say 20, i.e. outliers caused my measurement errors or instrumentation failures), will the results in Table 3.3 stay the same. Is ridge more robust than simple LS solutions?
5) For the remaining set of 30 samples, use the 4 models you obtained above to perform predictions. Compare the predicted outputs with the measured outputs given in the dataset. Can you get a sense which models give you good prediction in term of small prediction errors especially with outlier variables in training?