625.664 Computational Statistics

Problem Set 10

Associated Reading: Chapters 4 and 10 from Gentle

Complete the problems either by hand or using the computer and upload your final document to the Blackboard course site. All final submittals are to be in PDF form. Please document any code used to solve the problems and include it with your submission.

1. Show that the Fourier Trigonometric Family
$$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x, \ldots\}$$
is an orthogonal family over $0 \le x \le 2\pi$ with respect to the weight function $w(x) = 1$.

2. (a) Show that if $\{q_i(x)\}$ is a set of orthogonal functions, then it is a linearly independent set.
(b) Prove that the integrated mean squared error is the sum of the integrated variance and the integrated squared bias, that is,
$$IMSE(\hat{f}) = IV(\hat{f}) + ISB(\hat{f}).$$

3. Let $\{q_k : k = 1, \ldots, m\}$ be a set of orthogonal functions. Show that
$$\left\| \sum_{k=1}^{m} q_k \right\|^2 = \sum_{k=1}^{m} \|q_k\|^2 .$$
What is the common value of the expression above if the q_k are orthonormal? In these expressions, $\|\cdot\|$ represents an L_2 norm. Would a similar equation hold for a general L_p norm?

4. (a) Use the Gram-Schmidt orthogonalization process as described in Lecture 10C to derive the first four orthonormal Chebychev polynomials. Note that the range is $[-1, 1]$ and the weight function is $w(x) = (1 - x^2)^{1/2}$. (Please note that the set is to be orthonormal and not just orthogonal.)
(b) Posted on the course Blackboard is the data set Orthogonal.txt containing 1000 observations from a $N(0, .3^2)$ distribution (standard deviation of .3). Use the polynomials derived in part (a) to estimate the density f. Do you think that this is a good estimate of f? Be sure to explain why or why not.

5. (a) Which properties of a natural cubic spline does the following function posses and which does it not possess?
$$f(x) = \begin{cases} (x+1) + (x+1)^3 & x \in [-1, 0] \\ 4 + (x-1) + (x-1)^3 & x \in (0, 1] \end{cases}$$
(b) Find a natural cubic spline function whose knots are 1, 2, 3, and 4 that takes values $f(1) = 1$, $f(2) = 1/2$, $f(3) = 1/3$, and $f(4) = 1/4$.

6. (Extra Credit) Prove that the Fourier coefficients form the finite expansion in basis functions with the minimum mean squared error, that is prove

$$\left\| f - \sum_{k=0}^{j} c_k q_k \right\|^2 \leq \left\| f - \sum_{k=0}^{j} a_k q_k \right\|^2$$

where $\{c_k = \langle f, q_k \rangle\}$ are the Fourier coefficients and $\{a_k\}$ are any other constants. Assume that the basis set is orthonormal and note that $|| \cdot ||$ is the L_2 norm. Also, you may find it useful to use the results of Problem 3. (Hint: Write $\|f - a_0 q_0\|^2$ a function of a_0, $\langle f, f\rangle - 2a_0\langle f, q_0\rangle + a_0^2\langle q_0, q_0\rangle$, differentiate, set to zero for the minimum, and determine $a_0 = c_0$. This same approach can be done in multidimentions for $a_0, a_1, a_2, \ldots, a_k$, or else induction can be used from a_1 on.