# LINEAR REGRESSION MODELS W4315

## HOMEWORK 6 QUESTIONS

November 11, 2010

Due: 11/18/2010
Instructor: Frank Wood

**1. (15 points)** Refer to the design matrix given in "hw6p1.dat" on the course website. Read it into MATLAB ( the first **1** column is already added). Use "load hw6p1.dat" to load the file. What is the most complex model in terms of number of parameters that one could fit to this data?

Extra credit: If you fit a model with this number of parameters, how could you figure out which should be non-zero?

**2. (15 points)** Consider the classical matrix approach to multiple regression, i.e.

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{X}$ is a $n \times p$ design matrix whose first column is all 1's, $\epsilon \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{I}$ is an identity matrix. Prove the following:

a. The sum of squares error $SSE = \mathbf{e}'\mathbf{e}$ can be written in a matrix form:

$$SSE = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$$

b. We call the RHS of (2) a quadratic form. Prove that the matrix $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an idempotent matrix.

c. Prove that the rank of $\mathbf{A}$ defined in part (b) is $n - p$.

N.B. $p$ columns in design matrix means there are $p - 1$ predictors plus 1 intercept term. In your solutions please clearly notate the dimensions of all of the matrices.

**3. (45 points)** Suppose $X_1, ..., X_n$ are i.i.d. samples from $N(0, \sigma^2)$. Denote $\bar{X}$ as the sample mean. Prove $S = \sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \sigma^2\chi^2(n-1)$ following the steps below using Cochran's theorem:

a. Remember that we have the decomposition

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n\bar{X}^2$$

Show the matrices corresponding to all the three quadratic terms in (3).

b. Derive the rank of each matrix above.

c. Use Cochran's theorem to prove $S \sim \sigma^2\chi^2(n-1)$.