STATS 330

Advanced Statistical modeling

Mathematical details: Sheet 1

These notes are provided for those wanting a more mathematical description of some of the ideas we have been discussing. They are NOT examinable.

1. **Finding a power transformation to remove the funnel effect (see the left hand panel of Slide 26, Lecture 10)**

Suppose we transform a random variable Y using a power Y^p. How is the variance of Y^p related to the variance of Y? We can approximate Y^p by a Taylor expansion at the mean μ of Y: (this amounts to approximating the curve by the tangent at μ). We get

$$Y^p \approx \mu^p + (Y-\mu)\left.\frac{dy^p}{dy}\right|_{y=\mu}$$
$$= \mu^p + (Y-\mu)py^{p-1}$$

so taking the standard deviation of each side we get $s.d.(Y^p) \approx s.d.(Y)p\mu^{p-1}$. To make s.d.($Y^p$) independent of the mean, we should choose p so that $s.d.(Y)p\mu^{p-1}$ is approximately constant, or s.d.(Y) is proportional to μ^{1-p}. This is equivalent to (taking logs) log(sd(Y)) = C + (1-p)log(mean(Y)).

We can estimate p using the following procedure: take the responses in the regression and divide them onto say 5 groups on the basis of size. The responses in a group will have approximately the same mean, but the groups will have different means and sd's. Calculate the mean and sd of each group, and plot the log sd versus the log mean for the different groups. If the power transformation is going to work, the points should cluster about the line log(sd(Y)) = C + (1-p)log(mean(Y)). The slope of the line is 1-p, so we can work out p. This is exactly the plot in the left hand panel of slide 26.

2. **Hat matrix diagonals.**

The hat matrix diagonals (HMD's) are the diagonal elements of the matrix $H = X(X'X)^{-1}X'$, where X is the n x p matrix of explanatory variables (See slide 21 of lecture 5 for the definition of X.) Note that H is an nxn matrix, where n = number of observations. To work out the average HMD, we need to add up the diagonals and divide by n. The sum of the diagonal elements is called the trace of the matrix, and has the property that if A is an nxp matrix, and B is an pxn matrix, then the trace of the product AB is the same as the trace of the product BA. Applying this to H, we see that

Trace(H) = trace $(X(X'X)^{-1}X')$ = trace$(X' X(X'X)^{-1})$ = trace of a pxp identity matrix (all diagonal elements equal to1). This is just p. Thus, the average HMD is p/n.

The hat matrix has the property that H is symmetric (equal to its transpose) and idempotent (i.e. satisfies $H^2=H$. You can check this directly. Thus, if h_{ij} is the i,j element of H, we have

$$h_{ii} = \sum_{j=1}^{n} h_{ij}^2 \geq h_{ii}^2$$

This shows that $h_{ii} \geq 0$ and that $1 \geq h_{ii}$.

The fitted values are $\hat{y} = X\hat{\beta}$ where $\hat{\beta}$ is the vector of least squares estimates (the coefficients of the fitted plane). By slide 22 of lecture 5, these are the solutions to the equation $X'X\hat{\beta} = X'y$ or in matrix terms, $\hat{\beta} = (X'X)^{-1}X'y$. The hat matrix is derived by noting that $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy$.