Lab #2

Fall 2013 R. Beran

DUE 15 Oct 2013 in Discussion. Visit `www.stat.ucdavis.edu/~beran/s232a/vineyard.dat` for the data. Attach your computer code to your report.

The file `vineyard.dat` records the grape yield harvested in each row of a vineyard in three successive years. The first column gives the vineyard row number. Column $j+1$ reports the harvest yield in year j. We will analyze only the harvest data for year 3. This lab explores several competing computational algorithms for polynomial regression by least squares.

Let y_i denote the grape yield harvested from vineyard row i in year 3. The d-th degree polynomial model, with $0 \leq d \leq 51$, asserts that:

$$y_i = m_i + e_i, \qquad m_i = \sum_{r=0}^{d} \beta_r i^r, \qquad 1 \leq i \leq 52.$$

The $\{e_i\}$ are observation errors. The $\{\beta_r : 0 \leq r \leq d\}$ are unknown parameters. Note that the polynomial model of degree $d = 51$ is mathematically equivalent to the one-way layout model for this data, in which the $\{m_i\}$ are unrestricted.

a) Let $\beta = (\beta_0, \beta_1, \ldots \beta_d)'$ and let $\eta = (m_1, m_2, \ldots, m_{52})'$. Specify the matrix X such that $\eta = X\beta$. It is known that $\text{rank}(X) = d + 1$ when $0 \leq d \leq 51$. Thus, the matrix X is of full rank for this range of d.

Let $\hat{\eta}$ be the least squares estimator of η in the d-th degree polynomial model. Computing $\hat{\eta}$ accurately is challenging when d is not small. We consider **four competing** computational methods:

- The **classical solution** to the normal equation is $\hat{\eta} = X(X'X)^{-1}X'y$. Pertinent to the inversion is the R function `solve`.
- The **pseudoinverse** evaluation is $\hat{\eta} = XX^{+}y$. Pertinent to the pseudoinverse is the R function `ginv`, which becomes available after `library(MASS)`.
- The (reduced) **singular value decomposition** $X = ULV'$, discussed in class, implies that $\hat{\eta} = UU'y$. Pertinent to this decomposition is the R function `svd`.
- **Orthogonal polynomials** generate an orthogonal matrix W that has the same range space as X. Then $\hat{\eta} = WW'y$. Pertinent to the construction of orthogonal polynomials is the R function `poly`. Hint: Read the help file carefully.

b) Explain algebraically why $\hat{\eta} = UU'y = WW'y$.

c) For $d = 2$, use each of the four methods just described to compute $\hat{\eta}$. For each computation, plot the components of $\hat{\eta}$ against row number, interpolating linearly between the fitted points to guide the eye. Add the observed harvest data points for year 3 to the plot.

d) Repeat, to the extent possible, part c for polynomial degrees $d = 6$ and $d = 14$. Report which of the four methods for computing $\hat{\eta}$ breaks down. By referring to your experimental results, discuss the apparent reliability of these four methods for polynomial least squares fits.