Scientific Computing
Homework Eight

Brent Hyman

May 7, 2015

Exponential Linear Regression

→ Given a normal distribution pdf represented by the following:

$$q(t) = \frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2}$$

s.t.

$$Q_{\text{inf}} = \text{total amount of oil produced for all time}$$

$$\mu = \text{the mean, and in this case, is the maximum rate of oil production.}$$

$$\sigma = \text{the standard deviation}$$

Annual U.S. oil production measurements have been given. The idea in this assignment is to determine a best fit normal distribution curve to the data. In other words, σ and μ need to be calculated such that the normal distribution curve is a best fit along the given data.

Forming an exponential regression model for the data is very similar to forming a polynomial regression model. The exponential function just needs to be tweaked a little. The trick is to take the natural log of both sides.

$$ln(q(t)) = ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2})$$

$$= ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}) + ln(e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2})$$

$$= ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}) + \frac{-1}{2}(\frac{t-\mu}{\sigma})^2$$

$$= ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}) + \frac{-1}{2\sigma^2}(t-\mu)^2$$

$$= ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}) + \frac{-1}{2\sigma^2}(t^2 - 2\mu t + \mu^2)$$

$$= ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}) + \frac{-1}{2\sigma^2}t^2 + \frac{2\mu}{2\sigma^2}t - \frac{\mu^2}{2\sigma^2}$$

$$= \frac{-1}{2\sigma^2}t^2 + \frac{2\mu}{2\sigma^2}t + (\frac{-\mu^2}{2\sigma^2} + ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}))$$

In order to clean up this equation allow the following:

$$a = \frac{-1}{2\sigma^2}$$

$$b = \frac{2\mu}{2\sigma^2}$$

$$c = (\frac{-\mu^2}{2\sigma^2} + ln(\frac{Q_{\text{inf}}}{\sigma\sqrt{2\pi}}))$$

$$ln(q(t)) = z$$

so

$$z = at^2 + bt + c$$

This is a quadratic equation with parameters a, b, and c. The given OilProductionData.xls file holds the n pieces of data. The input for each piece of data is the time (in years) and the output is the oil production (barrels/year). In the case that I had only three pieces of data, the linear system would look like the following:

$$\begin{bmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

The solution of this linear system would yield parameters a, b, and c which would lead to defining the parameters σ and μ in such a way that would allow the normal curve to hit every data point exactly. But, the fact of the matter is that we have an over determined system to deal with:

$$\begin{bmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ t_n^2 & t_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ . \\ . \\ . \\ z_n \end{bmatrix}$$

This means that the best that can be done is to solve the least-squares problem in order to obtain parameters that make the curve fit the data with the least amount of error for all points. In other words, the parameter vector needs to be calculated s.t. the residual is minimized:

$$|| \begin{bmatrix} t_i^2 & t_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} - \begin{bmatrix} z_i \end{bmatrix} ||_{min}$$

This can be simplified by using A to represent the matrix, x for the parameter vector, and b to represent the right hand side vector:

$$||Ax - b||_{min}$$

In my USexponentialRegression script, I called matlab's built in QR function to aid in solving the least-squares problem. Then it uses the calculated parameters (contained in the least-squares soln) and the exponentialFunct function to form the best fit normal distribution curve for the data. The QR function returns a QR factorization of the A matrix and then the USexponentialRegression script solves the least squares problem by solving the following linear system:

$$Rx = Q^T b$$

The parameters a, b, and c turned out to be calculated as:

```
a:
 -7.0650e-004

b:
    2.7860

c:
 -2.7316e+003
```

The parameters σ and μ turned out to be calculated as:

```
sigma:
   26.6028

mu:
  1.9717e+003
```

The plot