Simple Linear Regression

Week 1

February 20, 2017

1 How to Fit a Line to Data

Given a plot of data a Linear Regression would be the linear function that minimizes the error between the points.

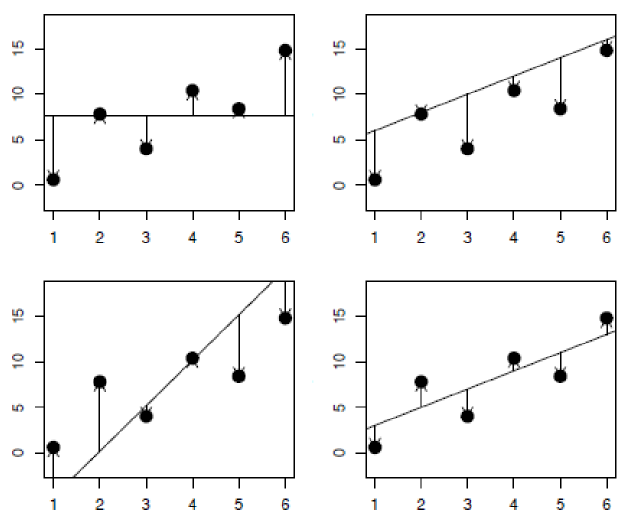

Figure 1: Differing error between linear functions

The Linear function would be of the form:

$$y = w_0 + w_1 \times x + \epsilon \tag{1}$$

The value ϵ represents the residual, the difference between the observed value and the value predicted by the model: $\epsilon_j = y_j - \hat{y}_j$.

To work out the linear function that best fits the points we use a concept called the *cost of a line* and it is the summed value of all the error for a line:

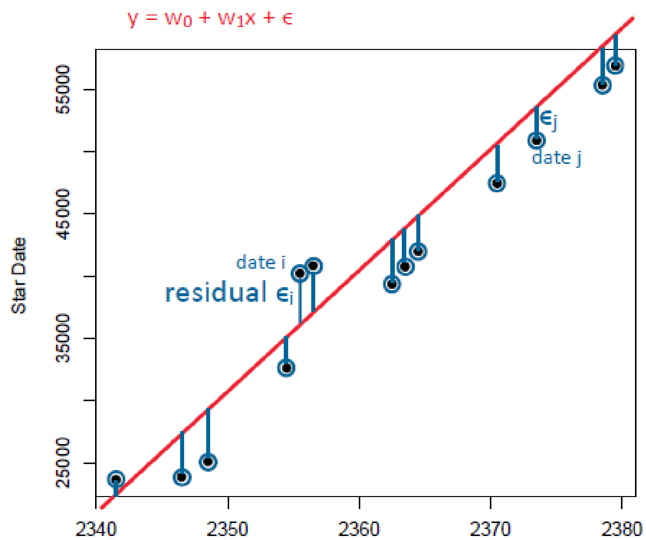

Figure 2: Differing error between linear functions

The value of the error would be ϵ_j, because ϵ can be positive or negative we take the square value such that we have positive numbers and sum them.
Thus the *Residual Sum of Squares* (***RSS***) is:

$$\sum_{i=1}^{n}[(\epsilon_i)^2] = \sum_{i=1}^{n}[(y_j - \hat{y_j})^2] \tag{2}$$

and the best fit for the line is defined as the one that minimimises the RSS.

2 How to Minimise the RSS

2.1 The Analytical Solution

Obviously the RSS will depend on the line chosen to fit the data, the line depends on two variables, the gradient and intersect, w_0 and w_1:

$$\begin{aligned} RSS &= \sum_{i=1}^{n}[(\epsilon_i)^2] \\ &= \sum_{i=1}^{n}\left[(y_i - \hat{y}_i)^2\right] \\ &= \sum_{i=1}^{n}\left[\left(y_i - w_0 - w_1 \times x_i\right)^2\right] \end{aligned} \tag{3}$$

Now given that the value y_i is a constant fixed value that is observed and x_i also refers to a fixed value observe that $RSS = f(x, y)$, this RSS value represents a 3-dimensional parabolic curve:

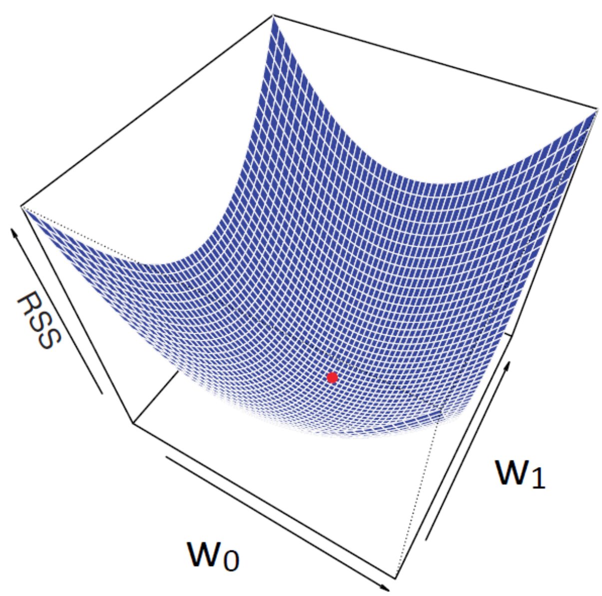

Figure 3: The variation of RSS given various values of w_0 and w_1

Thus we now know that the minimum RSS value will occur at a stationary point and this can be solved with calculus.

Example

$$RSS(w) = 10 - 8w + 2w^2 \tag{4}$$
$$\frac{dRSS}{dw} = 0 - 8 + 4w \tag{5}$$

Let the derivative equal zero then find the value of w where the RSS is a minimum value:

$$\begin{aligned} 0 &= -8 + 4w \\ 4w &= 8 \\ w &= 2 \end{aligned}$$

A quadratic function (like all convex functions) only has one minimum value and hence the solution is found. (A convex function is any function that a straight line could only cross twice, e.g. logarithmic is converse but Sine is not)

2.2 Least-squares Solution

Given that we now know that:

$$RSS = \sum_{i=1}^{n} \left[\left(y_i - w_0 - w_1 \times x_i\right)^2\right] \tag{6}$$

In order to find an equation that describes the minimum we will first find both the partial derivatives:

$$\begin{aligned}
\frac{\partial RSS}{\partial w_0} &= \frac{\partial}{\partial w_0}[\sum_{i=1}^{n} \left[\left(y_i - w_0 - w_1 \times x_i\right)^2\right]] \\
&= \sum_{i=1}^{n} \left[\frac{\partial}{\partial w_0}[\left(y_i - w_0 - w_1 \times x_i\right)^2\right]] \\
&= \sum_{i=1}^{n} \left[2 \times \left(y_i - w_0 - w_1 \times x_i\right)^1 \times -1\right] \\
&= -2\sum_{i=1}^{n} \left[\left(y_i - w_0 - w_1 x_i\right)^1\right]
\end{aligned} \tag{7}$$

$$\begin{aligned}
\frac{\partial RSS}{\partial w_1} &= \frac{\partial}{\partial w_1}[\sum_{i=1}^{n} \left[\left(y_i - w_0 - w_1 \times x_i\right)^2\right]] \\
&= \sum_{i=1}^{n} \left[\frac{\partial}{\partial w_1}[\left(y_i - w_0 - w_1 \times x_i\right)^2\right]] \\
&= \sum_{i=1}^{n} \left[2 \times \left(y_i - w_0 - w_1 \times x_i\right)^1 \times x_i\right] \\
&= 2 \times \sum_{i=1}^{n} \left[x_i \times \left(y_i - w_0 - w_1 x_i\right)^1\right]
\end{aligned} \tag{8}$$

Now we will let both partial derivatives equal zero, the point at which the RSS is minimized:

$$
\begin{aligned}
0 &= \frac{\partial RSS}{\partial w_0} && (9)\\
0 &= -2\sum_{i=1}^{n}\left[\big(y_i - w_0 - w_1x_i\big)^1\right] \\
&= \sum_{i=1}^{n}\left[\big(y_i - w_0 - w_1x_i\big)^1\right] \\
&= \sum_{i=1}^{n}[y_i] - \sum_{i=1}^{n}[w_0] + \sum_{i=1}^{n}[w_1x_i] \\
&= \sum_{i=1}^{n}[y_i] - nw_0 + w_1\sum_{i=1}^{n}[x_i] \\
nw_0 &= \sum_{i=1}^{n}[y_i] + w_1\sum_{i=1}^{n}[x_i] \\
w_0 &= \frac{\sum_{i=1}^{n}[y_i]}{n} + \frac{w_1\sum_{i=1}^{n}[x_i]}{n} && (10)
\end{aligned}
$$

Observe that in (15) the left term is the average value of y_i and the right term is the average value of x_i.

$$
\begin{aligned}
0 &= \frac{\partial RSS}{\partial w_1} && (11)\\
&= 2\times\sum_{i=1}^{n}\left[x_i\times\big(y_i - w_0 - w_1x_i\big)\right] \\
&= 2\times\sum_{i=1}^{n}\left[x_i\times y_i - w_0x_i + w_1(x_i)^2\right] \\
&= \sum_{i=1}^{n}[x_iy_i] - \sum_{i=1}^{n}[w_0x_i] - \sum_{i=1}^{n}[w_1(x_i)^2] \\
w_1\sum_{i=1}^{n}[(x_i)^2] &= \sum_{i=1}^{n}[x_iy_i] - w_0\sum_{i=1}^{n}[x_i] && (12)
\end{aligned}
$$

Now we will substitute (15) into (12) such that the equation is only in terms of x_i and y_i.

$$
\begin{aligned}
w_1 \sum_{i=1}^{n}[(x_i)^2] &= \sum_{i=1}^{n}[x_i y_i] - [\frac{\sum_{i=1}^{n}[y_i]}{n} + \frac{w_1 \sum_{i=1}^{n}[x_i]}{n}] \sum_{i=1}^{n}[x_i] \\
&= \sum_{i=1}^{n}[x_i y_i] - \frac{\sum_{i=1}^{n}[x_i] \sum_{i=1}^{n}[y_i]}{n} - w_1 \frac{(\sum_{i=1}^{n}[x_i])^2}{n} \\
w_1 \sum_{i=1}^{n}[(x_i)^2] + w_1 \frac{(\sum_{i=1}^{n}[x_i])^2}{n} &= \sum_{i=1}^{n}[x_i y_i] - \frac{\sum_{i=1}^{n}[x_i] \sum_{i=1}^{n}[y_i]}{n} \\
w_1 [\sum_{i=1}^{n}[(x_i)^2] + \frac{(\sum_{i=1}^{n}[x_i])^2}{n}] &= \sum_{i=1}^{n}[x_i y_i] - \frac{\sum_{i=1}^{n}[x_i] \sum_{i=1}^{n}[y_i]}{n} \\
w_1 &= \frac{\sum_{i=1}^{n}[x_i y_i]}{[\sum_{i=1}^{n}[(x_i)^2] + \frac{(\sum_{i=1}^{n}[x_i])^2}{n}]} - \frac{\frac{\sum_{i=1}^{n}[x_i] \sum_{i=1}^{n}[y_i]}{n}}{[\sum_{i=1}^{n}[(x_i)^2] + \frac{(\sum_{i=1}^{n}[x_i])^2}{n}]} \\
w_1 &= \frac{\sum_{i=1}^{n}[x_i y_i] - \frac{\sum_{n}^{i=1}[x_i] \sum_{n}^{i=1}[y_i]}{n}}{[\sum_{i=1}^{n}[(x_i)^2] + \frac{(\sum_{i=1}^{n}[x_i])^2}{n}]} \qquad (13)
\end{aligned}
$$

3 Conclusion

A Linear Regression is a linear function that has the lowest Sum of Squared Error between the function values and the observed values and is given by the equation:

$$y = w_0 + w_1 x + \epsilon \qquad (14)$$

Such that the coefficients are provided by the equations:

$$
\begin{aligned}
w_0 &= \frac{\sum_{i=1}^{n}[y_i]}{n} + \frac{w_1 \sum_{i=1}^{n}[x_i]}{n} \qquad (15) \\
w_1 &= \frac{\sum_{i=1}^{n}[x_i y_i] - \frac{\sum_{n}^{i=1}[x_i] \sum_{n}^{i=1}[y_i]}{n}}{[\sum_{i=1}^{n}[(x_i)^2] + \frac{(\sum_{i=1}^{n}[x_i])^2}{n}]} \qquad (16)
\end{aligned}
$$