Sum of Squared Errors - Sanity Check

INM431 week 5 - daniel.sikar@city.ac.uk + big sister

Given a line equation:

$$y = b + mx$$

where b is the intercept, and m is the slope. We change this into the form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{1}$$

and define an error function:

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

where y_i is the ith observation of y, and corresponds to the input x_i, and $\hat{y}_i$ is the corresponding prediction. Substituting (1) in (2):

$$E = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

We want to minimize the error E and can see it is a function of $\hat{\beta}_0$ and $\hat{\beta}_1$ so give it a more descriptive name:

$$SSE(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{3}$$

We take two derivatives, one with respect to $\hat{\beta}_0$ and the other with respect to $\hat{\beta}_1$. Since our error function consists of two functions, one outside function (exponentiation) and one inside function, we apply the chain rule for a general case:

$$g(f(x)) = \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial x}$$

For $\hat{\beta}_0$ we have:

$$\frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial \hat{\beta}_0} = \frac{\partial g}{\partial f} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \cdot \frac{\partial f}{\partial \hat{\beta}_0} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

We apply the derivative sum rule - the derivative of the sum is equal to the sum of the derivatives:

$$\begin{aligned}
\frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial \hat{\beta}_0} &= \sum_{i=1}^{n} \frac{\partial g}{\partial f} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \cdot \frac{\partial f}{\partial \hat{\beta}_0} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \cdot (-1) \\
&= -2 \sum_{i=1}^{n} y_i + 2 \sum_{i=1}^{n} \hat{\beta}_0 + 2 \sum_{i=1}^{n} \hat{\beta}_1 x_i
\end{aligned}$$

Since we are trying to minimize:

$$\frac{\partial}{\partial \hat{\beta}_0} SSE(\hat{\beta}_0, \hat{\beta}_1)$$

we set the result to zero and solve for $\hat{\beta}_0$:

$$-2\sum_{i=1}^{n} y_i + 2\sum_{i=1}^{n} \hat{\beta}_0 + 2\sum_{i=1}^{n} \hat{\beta}_1 x_i = 0$$

$$\begin{aligned}
2\sum_{i=1}^{n} \hat{\beta}_0 &= 2\sum_{i=1}^{n} y_i - 2\sum_{i=1}^{n} \hat{\beta}_1 x_i \\
\sum_{i=1}^{n} y_i = n\bar{y} & \\
2\sum_{i=1}^{n} \hat{\beta}_0 &= 2n\bar{y} - 2\sum_{i=1}^{n} \hat{\beta}_1 x_i \\
2\sum_{i=1}^{n} \hat{\beta}_0 &= 2n\bar{y} - \hat{\beta}_1 2\sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} x_i = n\bar{x} & \\
2\sum_{i=1}^{n} \hat{\beta}_0 &= 2n\bar{y} - \hat{\beta}_1 2n\bar{x} \\
2n\hat{\beta}_0 &= 2n\bar{y} - \hat{\beta}_1 2n\bar{x}
\end{aligned}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{4}$$

For $\hat{\beta}_1$ we have:

$$\begin{aligned}
\frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial \hat{\beta}_1} &= \sum_{i=1}^{n} \frac{\partial g}{\partial f}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \cdot \frac{\partial f}{\partial \hat{\beta}_1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= 2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \cdot (-x_i) \\
&= -2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} \hat{\beta}_0 x_i + 2\sum_{i=1}^{n} \hat{\beta}_1 x_i^2
\end{aligned}$$

Setting result equal to zero:

$$
\begin{aligned}
-2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} \hat{\beta}_0 x_i + 2\sum_{i=1}^{n} \hat{\beta}_1 x_i^2 &= 0 \\
-\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} \hat{\beta}_0 x_i + \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 &= 0 \\
-\sum_{i=1}^{n} x_i y_i + \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 &= 0
\end{aligned}
$$

Substituting $\hat{\beta}_0$ from (4):

$$
\begin{gathered}
-\sum_{i=1}^{n} x_i y_i + (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \\
-\sum_{i=1}^{n} x_i y_i + \bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \\
-\sum_{i=1}^{n} x_i y_i + \bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 (\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2) = 0 \\
\hat{\beta}_1 (\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2) = -\sum_{i=1}^{n} x_i y_i + \bar{y} \sum_{i=1}^{n} x_i \\
\hat{\beta}_1 (\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2) = \bar{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i
\end{gathered}
$$

$$\hat{\beta}_1 = \frac{\bar{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i}{\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

$$\sum_{i-1}^{n} x_i = n\bar{x}$$

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\bar{y} n \bar{x} - \sum_{i=1}^{n} x_i y_i}{\bar{x} n \bar{x} - \sum_{i=1}^{n} x_i^2} \\
\hat{\beta}_1 &= \frac{n \bar{y} \bar{x} - \sum_{i=1}^{n} x_i y_i}{n \bar{x}^2 - \sum_{i=1}^{n} x_i^2} \\
\hat{\beta}_1 &= \frac{n \bar{y} \bar{x} - \sum_{i=1}^{n} x_i y_i}{n \bar{x}^2 - \sum_{i=1}^{n} x_i^2} \cdot \frac{-1}{-1}
\end{aligned}
$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2} \tag{5}$$

Removing n term from nominator in (5):

$$
\begin{aligned}
\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} &= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\
n\bar{x}\bar{y} = \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y} \\
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \\
&= \sum_{i=1}^{n} x_i y_i - n\frac{\sum_{i=1}^{n} x_i}{n}\bar{y} - n\bar{x}\bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - n\bar{x}\bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - \bar{x}n\bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y} \\
\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - \bar{x}n\frac{\sum_{i=1}^{n} y_i}{n} + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - \bar{x}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - \sum_{i=1}^{n} \bar{x}y_i + \sum_{i=1}^{n} \bar{x}\bar{y} \\
&= \sum_{i=1}^{n} \big(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}\big) \\
&= \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\end{aligned}
$$

Substituting in (5):

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{6}
$$

Removing n term from denominator in (6):

$$\begin{aligned}
\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 &= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\
n\bar{x} = \sum_{i=1}^{n} x_i \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\
&= \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2
\end{aligned}$$

Substituting in (6):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{7}$$

Solving for example given in class:

```
x = [2 4 10] % average speed on freeway
y = [75 45 35] % patrol cars deployed

xbar = (2 + 4 + 10) / 3 % 5.33333333333
ybar = (75 + 45 + 35) / 3 % 51.6666666667

// beta hat subscript 1 ~ slope
m = ((2 - xbar)(75 - ybar) + ...
     (4 - xbar)(45 - ybar) + ...
     (10 - xbar)(35 - ybar)) / ...
     ((2 - xbar)^2 + (4 - xbar)^2 + (10 - xbar)^2) % -4.23076923077

// beta hat subscript 0 ~ intercept
b = ybar - -4.23076923077 * xbar % 74.2307692308

% line of best fit
% y = 74.2307692308 - 4.23076923077 * x
```

```
% prediction y for x = 5
% y = 53.0769230769mph

% MATLAB solution for b and m
glmfit([2 4 10], [75 45 35]) % coefficient estimates
 ans =

   74.2308
   -4.2308
```

Second Partial Derivative Test

To find out if $\hat{\beta}_0$ and $\hat{\beta}_1$ is a local minimum, local maximum, a saddle point or cannot be determined, we need to compute the second order partial derivatives, and the mixed second order partial derivative of the error function with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. The point $(\hat{\beta}_0, \hat{\beta}_1)$ is a local minimum if:

$$\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} \right)^2 > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial \hat{\beta}_0^2} > 0$$

The point $(\hat{\beta}_0, \hat{\beta}_1)$ is a local maximum if:

$$\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} \right)^2 > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial \hat{\beta}_0^2} < 0$$

The point $(\hat{\beta}_0, \hat{\beta}_1)$ is a saddle if:

$$\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} \right)^2 < 0$$

The Second Derivative Test is indeterminate if:

$$\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} \right)^2 = 0$$

Returning to the forms:

$$
\begin{gathered}
\frac{\partial f}{\partial \hat{\beta}_0} SSE(\hat{\beta}_0, \hat{\beta}_1) = 0 \\
-2\sum_{i=1}^{n} y_i + 2\sum_{i=1}^{n} \hat{\beta}_0 + 2\sum_{i=1}^{n} \hat{\beta}_1 x_i = 0 \\
\frac{\partial f}{\partial \hat{\beta}_1} SSE(\hat{\beta}_0, \hat{\beta}_1) = 0 \\
-2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} \hat{\beta}_0 x_i + 2\sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0
\end{gathered}
$$

we obtain the second order partial derivatives:

$$
\begin{gathered}
\frac{\partial^2 f}{\partial \hat{\beta}_0^2} = 2n \\
\frac{\partial^2 f}{\partial \hat{\beta}_1^2} = 2\sum_{i=1}^{n} x_i^2
\end{gathered}
$$

and the mixed second order partial derivative:

$$
\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} = 2\sum_{i=1}^{n} x_i
$$

$$
\begin{aligned}
\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1}\right)^2 &= (2n) \cdot \left(2\sum_{i=1}^{n} x_i^2\right) - \left(2\sum_{i=1}^{n} x_i\right)^2 \\
&= (2n) \cdot \left(2\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i\right) - \left(2\sum_{i=1}^{n} x_i\right)^2 \\
&= (2n) \cdot \left(2\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i\right) - \left(2\sum_{i=1}^{n} x_i 2\sum_{i=1}^{n} x_i\right) \\
&= (2n) \cdot (2n\bar{x}n\bar{x}) - (2n\bar{x}2n\bar{x}) \\
&= (2n) \cdot (2n\bar{x}n\bar{x}) - (4n^2\bar{x}^2) \\
&= (2n) \cdot (2n^2\bar{x}^2) - (4n^2\bar{x}^2) \\
&= 4n^3\bar{x}^2 - 4n^2\bar{x}^2
\end{aligned}
$$

Since:

$$
\begin{gathered}
n > 0 \\
4n^3\bar{x}^2 > 4n^2\bar{x}^2 \\
\frac{\partial^2 f}{\partial \hat{\beta}_0^2} \cdot \frac{\partial^2 f}{\partial \hat{\beta}_1^2} - \left(\frac{\partial^2 f}{\partial \hat{\beta}_0 \hat{\beta}_1} \right)^2 > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial \hat{\beta}_0^2} > 0
\end{gathered}
$$

the point $(\hat{\beta}_0, \hat{\beta}_1)$ is a local minimum.