Lab Solutions 04

1. Write the Equation as an inner product. Expand and distribute the terms so that you have a loss function written a liner combination of matrix products.

The loss function can be expanded as:

$$
\begin{align}
\mathcal{L} &= ||y - X\beta||_2^2 \tag{4.1}\\
&= (y - X\beta)^t(y - X\beta) \tag{4.2}\\
&= (y^t - (X\beta)^t)(y - X\beta) \tag{4.3}\\
&= (y^t - \beta^t X^t)(y - X\beta) \tag{4.4}\\
&= y^t y + \beta^t X^t X\beta - \beta^t X^t y - y^t X\beta \tag{4.5}
\end{align}
$$

Notice that all of these terms are scalar quantities, so we can replace the value $\beta^t X^t y$ with its transpose $y^t X\beta$. This simplifies the formula to:

$$\mathcal{L} = y^t y + \beta^t X^t X\beta - 2y^t X\beta. \tag{4.6}$$

2. Convince yourself that the matrix $X^t X$ is equal to its own inverse.

This is true just because of the rules for distributing the transpose operator:

$$(X^t X)^t = X^t (X^t)^t = X^t X. \tag{4.7}$$

3. Use the gradient rules we had last time to compute the gradient of the loss function for linear regression.

The gradient is given, using the results from last time, as:

$$\nabla_\beta \mathcal{L} = 2X^t X\beta - 2X^t y. \tag{4.8}$$

We already did the hard part last time.

4. Set the gradient equal to zero. The result is known as the *normal equations*. Isolate β on one side using the matrix inverse.

Setting the gradient equal to zero yields:

$$2X^t X\beta = 2X^t y. \tag{4.9}$$

And solving for β gives:

$$\beta = (X^t X)^{-1} X^t y. \tag{4.10}$$

Assuming, of course, that the inverse exists.

5. Now, with a known quantity for β, write down an equation for $\widehat{y}$. This should take the form:

$$\widehat{y} = Hy \tag{4.11}$$

For some matrix H. The matrix here is called the "hat" matrix because it puts a hat on the quantities of y. Show that $H^2 = HH = H$.

The equation for $\widehat{y}$ is given by:

$$\begin{aligned} \widehat{y} &= X\beta && (4.12) \\ &= X(X^t X)^{-1} X^t y. && (4.13) \end{aligned}$$

So that the hat matrix is:

$$H = X(X^t X)^{-1} X^t. \tag{4.14}$$

Showing that $H^2 = H$ is just a matter of expanding and cancelling terms:

$$\begin{aligned} H^2 &= X(X^t X)^{-1} X^t X (X^t X)^{-1} X^t && (4.15) \\ &= X(X^t X)^{-1} (X^t X)(X^t X)^{-1} X^t && (4.16) \\ &= X(X^t X)^{-1} X^t && (4.17) \\ &= H. && (4.18) \end{aligned}$$

6. Assume that the "true" value of y is given by:

$$y = Xb + \epsilon \tag{4.19}$$

For some vector of random errors ϵ. Plug this into your equation for β and show that β can be written as b plus another term that should be small if the noise terms are small.

Here we have:

$$\begin{aligned} \beta &= (X^t X)^{-1} X^t y && (4.20) \\ &= (X^t X)^{-1} X^t (Xb + \epsilon && (4.21) \\ &= (X^t X)^{-1} X^t X b + (X^t X)^{-1} X^t \epsilon && (4.22) \\ &= b + (X^t X)^{-1} X^t \epsilon. && (4.23) \end{aligned}$$

7. Show that the residuals ($y - X\beta$) are perpendicular to the fitted values $\widehat{y}$. That is, show that the dot product between the two is zero. (Note: you can use the fact that $(X^tX)^{-1}$ is equal to its own transpose).

Notice that:

$$\begin{aligned}
\beta^t &= \left((X^tX)^{-1}X^ty\right)^t && (4.24)\\
&= y^tX((X^tX)^{-1})^t && (4.25)\\
&= y^tX(X^tX)^{-1} && (4.26)
\end{aligned}$$

Then, plugging in, everything eventually cancels:

$$\begin{aligned}
(y - X\beta)^t(X\beta) &= y^tX\beta - \beta^tX^tX\beta && (4.27)\\
&= y^tX(X^tX)^{-1}X^ty - X(X^tX)^{-1}X^tX(X^tX)^{-1}X^t && (4.28)\\
&= y^tX(X^tX)^{-1}X^ty - X(X^tX)^{-1}X^t && (4.29)\\
&= 0 && (4.30)
\end{aligned}$$

8. Continue by following the notes in the `class04.Rmd`