# 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:

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

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

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

**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:

$$(X^t X)^t = X^t (X^t)^t = X^t X. \qquad (4.7)$$

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

Assuming you have already made the substitution (see solution to question 1 for justification):

$$y^t X\beta = \beta^t X^t y \qquad (4.8)$$

The gradient is given by using the results from last time to each component:

$$\begin{aligned}
\nabla_\beta \mathcal{L} &= \nabla_\beta \left[y^t y + \beta^t X^t X\beta - 2\beta^t X^t y\right] && (4.9)\\
&= 2X^t X\beta - 2X^t y. && (4.10)
\end{aligned}$$

Notice that we can apply the rule for the gradient of $b^t Ab$ by setting $A = (X^t X)$.

**4. Set the gradient equal to zero. The result is known as the *normal equations*.**

**Isolate $\beta$ on one side using the matrix inverse.**

Setting the gradient equal to zero yields:

$$2X^tX\beta = 2X^ty. \tag{4.11}$$

And solving for $\beta$ gives:

$$\beta = (X^tX)^{-1}X^ty. \tag{4.12}$$

Assuming, of course, that the inverse exists.

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

$$\widehat{y} = Hy \tag{4.13}$$

**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:

$$\begin{aligned} \widehat{y} &= X\beta && (4.14)\\ &= X(X^tX)^{-1}X^ty. && (4.15) \end{aligned}$$

So that the hat matrix is:

$$H = X(X^tX)^{-1}X^t. \tag{4.16}$$

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

$$\begin{aligned} H^2 &= X(X^tX)^{-1}X^tX(X^tX)^{-1}X^t && (4.17)\\ &= X(X^tX)^{-1}(X^tX)(X^tX)^{-1}X^t && (4.18)\\ &= X(X^tX)^{-1}X^t && (4.19)\\ &= H. && (4.20) \end{aligned}$$

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

$$y = Xb + \epsilon \tag{4.21}$$

**For some vector of random errors $\epsilon$. Plug this into your equation for $\beta$ and show that $\beta$ can be written as $b$ plus another term that should be small if the noise terms**

**are small.**

Here we have:

$$\beta = (X^tX)^{-1}X^ty \tag{4.22}$$
$$= (X^tX)^{-1}X^t(Xb+\epsilon \tag{4.23}$$
$$= (X^tX)^{-1}X^tXb + (X^tX)^{-1}X^t\epsilon \tag{4.24}$$
$$= b + (X^tX)^{-1}X^t\epsilon. \tag{4.25}$$

**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:

$$\beta^t = \left((X^tX)^{-1}X^ty\right)^t \tag{4.26}$$
$$= y^tX((X^tX)^{-1})^t \tag{4.27}$$
$$= y^tX(X^tX)^{-1} \tag{4.28}$$

Then, plugging in, everything eventually cancels:

$$(y - X\beta)^t(X\beta) = y^tX\beta - \beta^tX^tX\beta \tag{4.29}$$
$$= y^tX(X^tX)^{-1}X^ty - X(X^tX)^{-1}X^tX(X^tX)^{-1}X^t \tag{4.30}$$
$$= y^tX(X^tX)^{-1}X^ty - X(X^tX)^{-1}X^t \tag{4.31}$$
$$= 0 \tag{4.32}$$

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