Jared Yu

Module 3 Discussion Reply

I think overall, we had the same approach. From my understanding we both tried to derive the equations for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$, then compared them to see if they were the same. In your approach, I see that your solved for them by optimizing with the least-squares formula $S(\widehat{\boldsymbol{\beta}})$ and $S(\widehat{\boldsymbol{\alpha}})$. I tried a similar approach originally but was getting lost in the process. During office hours, the professor mentioned that I could instead try the matrix formulation for the normal equations. I got stuck there too, but after more time trying to figure out the inverse, I think that I may have found the correct formula.

In my discussion, I try to show basically that the formula is invariant to something like centering the values with the sample mean. I see that you have a similar result, where you indicate that the intercepts $\hat{\beta}_0 \neq \hat{\alpha}_0$, which is something that I noted also. I think however that the other parameters match, i.e., $\hat{\beta}_1 = \hat{\alpha}_1$ and $\hat{\beta}_2 = \hat{\alpha}_2$. This seems to be noted also in the supplemental lecture for module 3 and so I suppose it is not necessary to state in our discussion.

Regarding the random versus non-random regressors, the issue that I had brought up is that since my derivation is based on the matrix formulation that requires that the inverse (i.e., $(\mathbf{X}'\mathbf{X})^{-1}$) exists. It is possible that since there are some values that are in the denominator, if those evaluate to 0 based on the sample data, then the parameters would not be possible to derive. Specifically, it seemed to require that the sample correlation not be equal to -1 or 1. Given fixed data, this can be determined easily or even ensured if it comes from simulation. However, if the data is randomly generated, then it is not possible to ensure until it is checked.

I like what you did with the simulated data. It is very interesting to see through simulation that the slope of the two models look to be identical, but the y-intercept is different. It seems then that with random versus fixed data, there will still be a difference in the two model parameters.