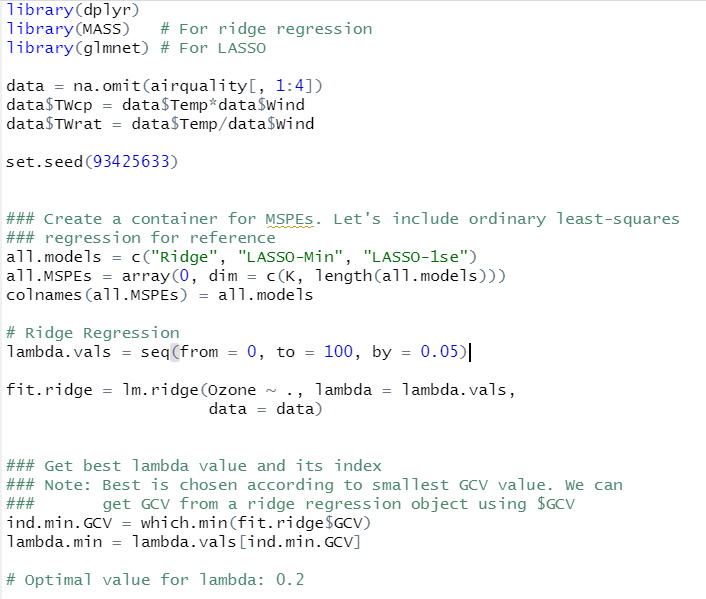
**Application**

Refer to the Air Quality data described previously, and the analyses we have done with Ozone as the response variable, and the five explanatory variables (including the two engineered features).

1. Use ridge regression on the data:

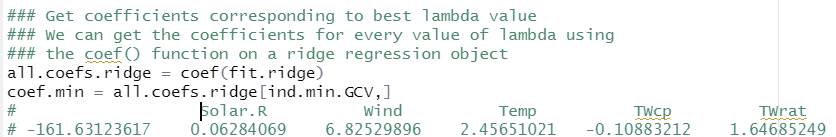
(a) Using GCV, identify the optimal *\_* from the sequence 0 to 100 by 0.05. **Report**

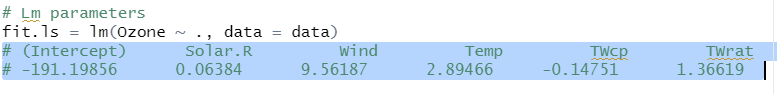
**the optimal value for** *lambda***.**



(b) **Report the parameter estimates for the optimal model, and compare**

**them to those from least squares. Are the ridge estimates all smaller.**

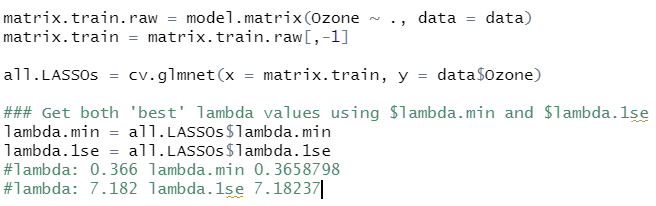




No, not all ridge estimates smaller.

2. Use LASSO on the data

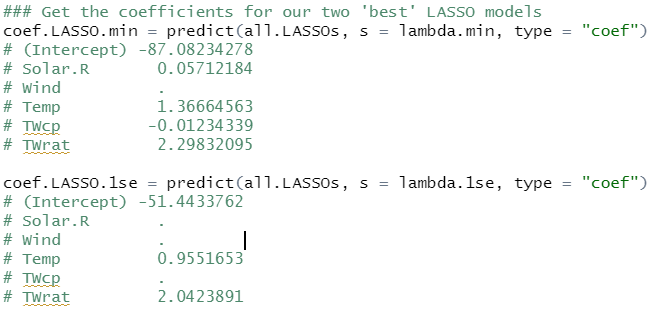
(a) Using default CV in cv.glmnet, identify the optimal *lambda*, *lambdamin*, and the “1SE” *lambda*, *lambda*1*SE*. **Report both numbers.**



(b) **For both selections of** *\_***, report the parameter estimates for the optimal**

**model. Comment on how the two sets of estimates differ from each**

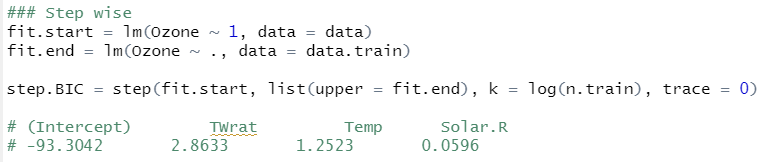
**other.**



LASSO.min drops Wind variable and LASSO.1se drops Solar.R, Wind, and TWcp

(c) Compare the variables selected by both versions of LASSO to those from the

hybrid stepwise from Lecture 5. **What differences are there?**



Lasso.min selects Solar.R, Temp, Twcp, and TWrat as variables.

Lasso.1se selects Temp and TWrat as variables

Hybrid stepwise selects TWrat, Temp, and Solar.R as variables.

Lasso.1se seems to be the simplest one.