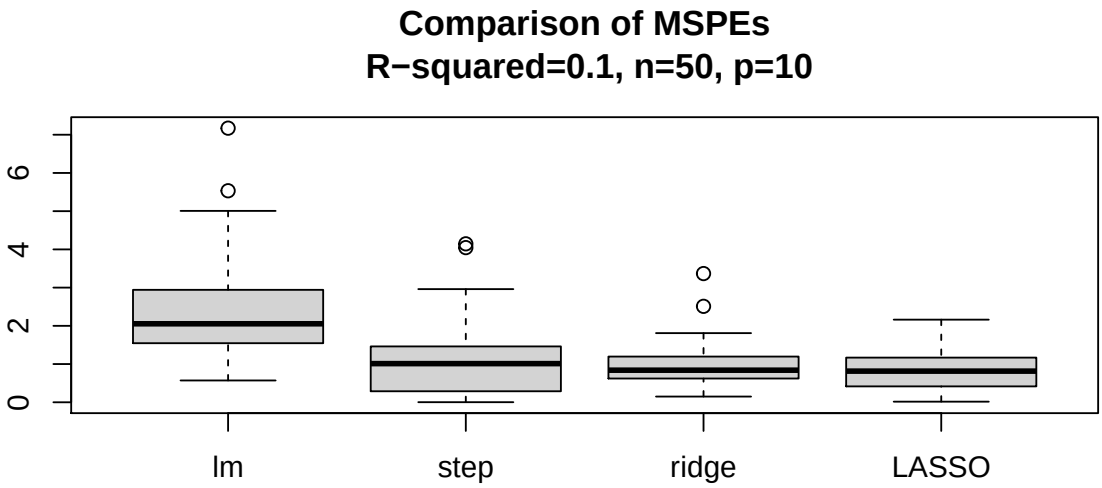
STAT452/652 Solution to HW05 - Lecture 6

Due on Oct 16, 2020

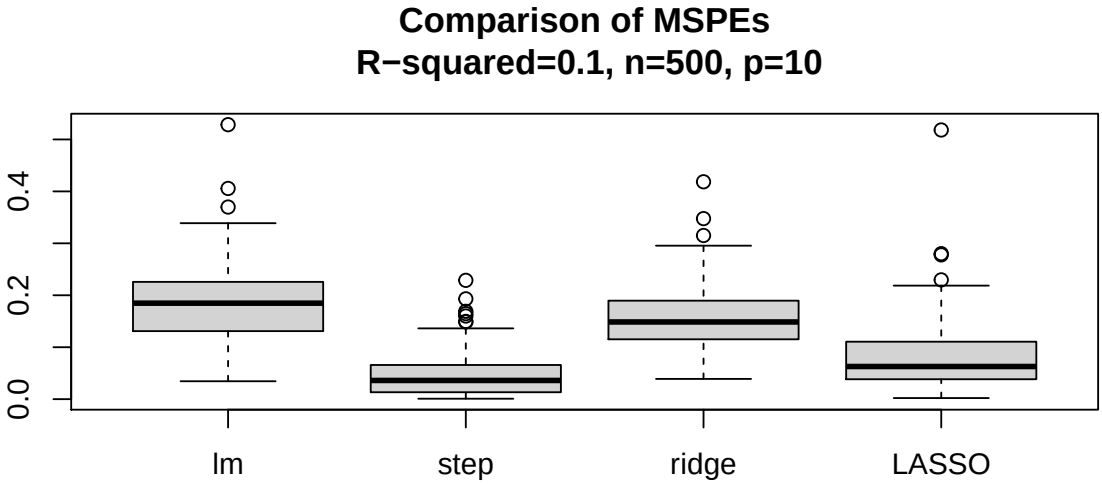
1 Concepts

1.1 Question 1



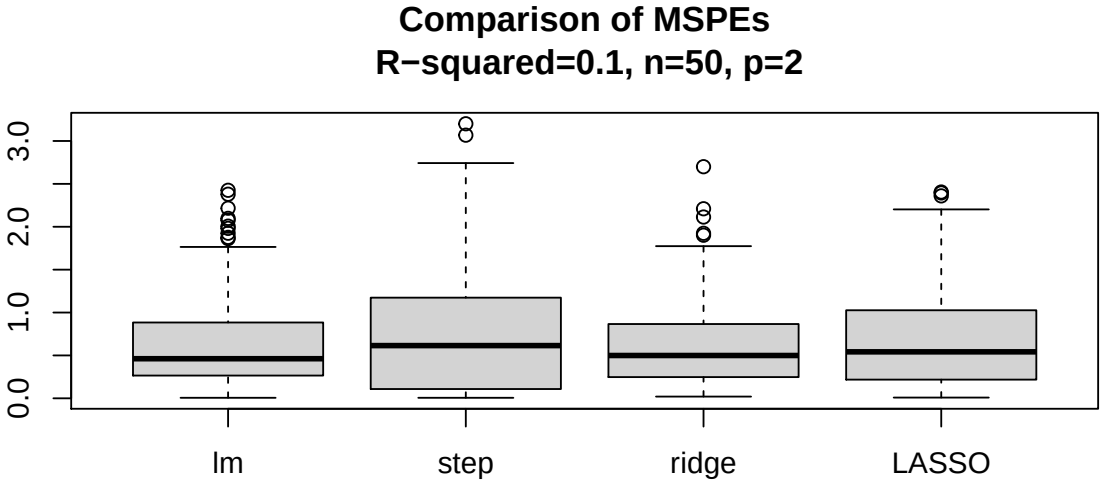
For 50/10/3, LASSO looks like the best model, although ridge is pretty similar, while stepwise is close.

1.2 Question 2



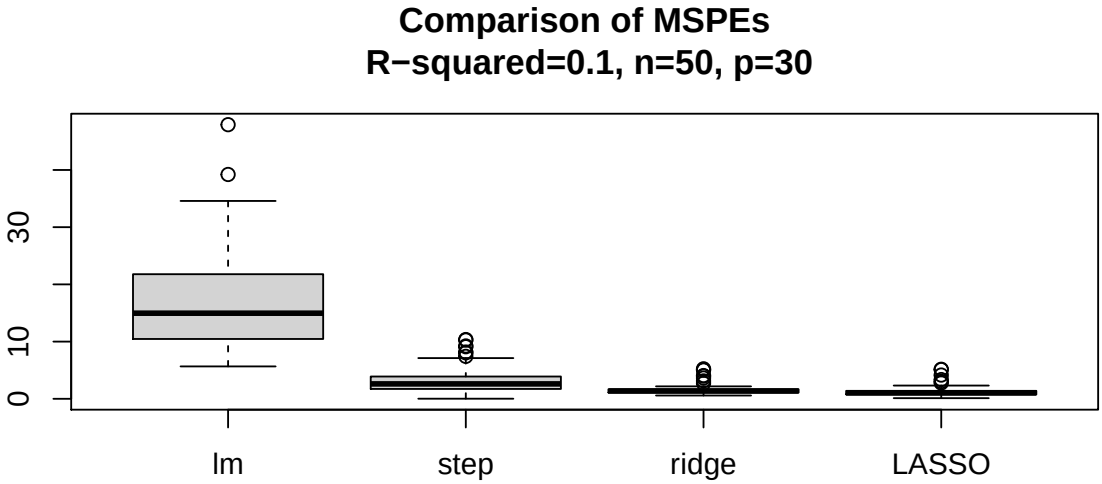
For 500/10/3, stepwise looks like the best model. LASSO is close.

1.3 Question 3



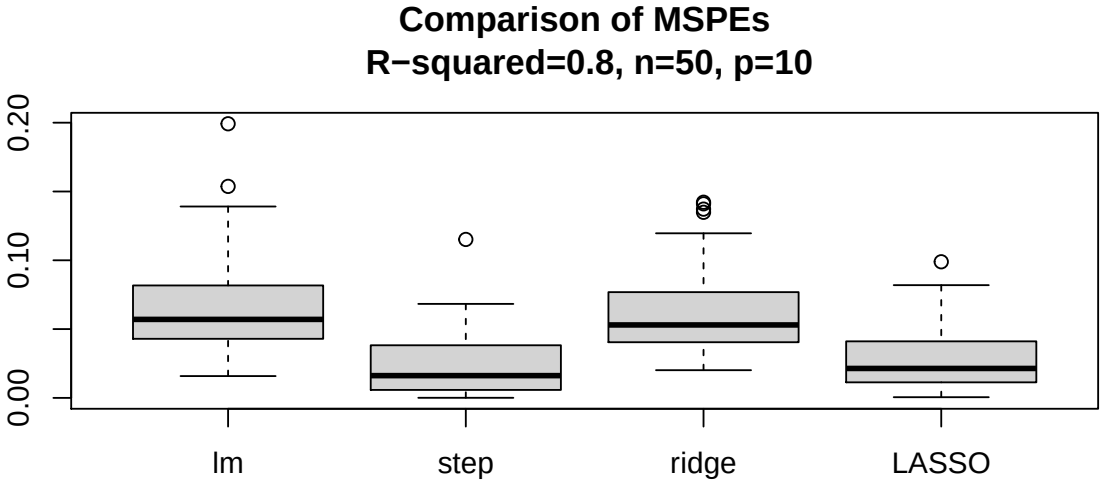
For 50/2/3, all methods are pretty similar, but LM and ridge seem a little better with fewer high values.

1.4 Question 4



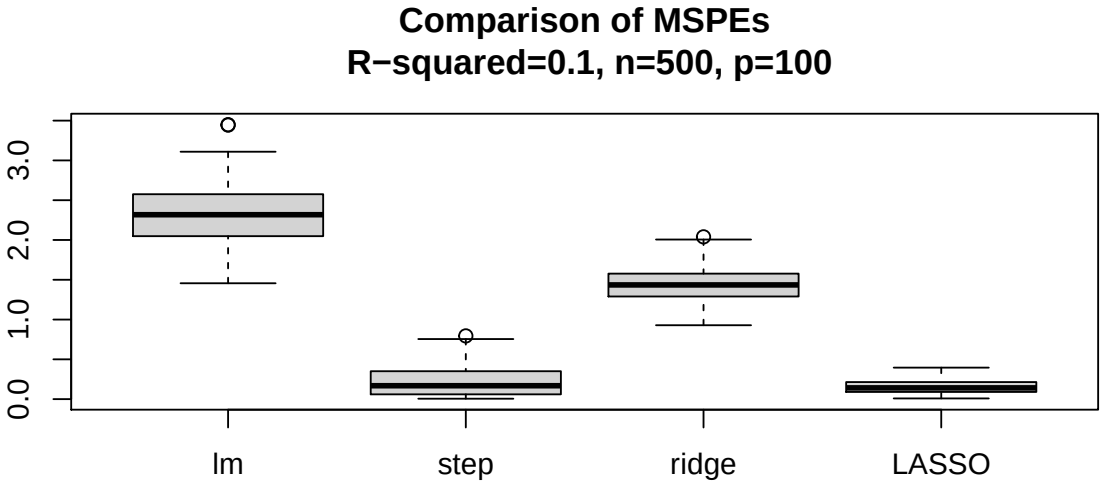
For 50/30/3, LASSOs look like the best model, but ridge is close.

1.5 Question 5



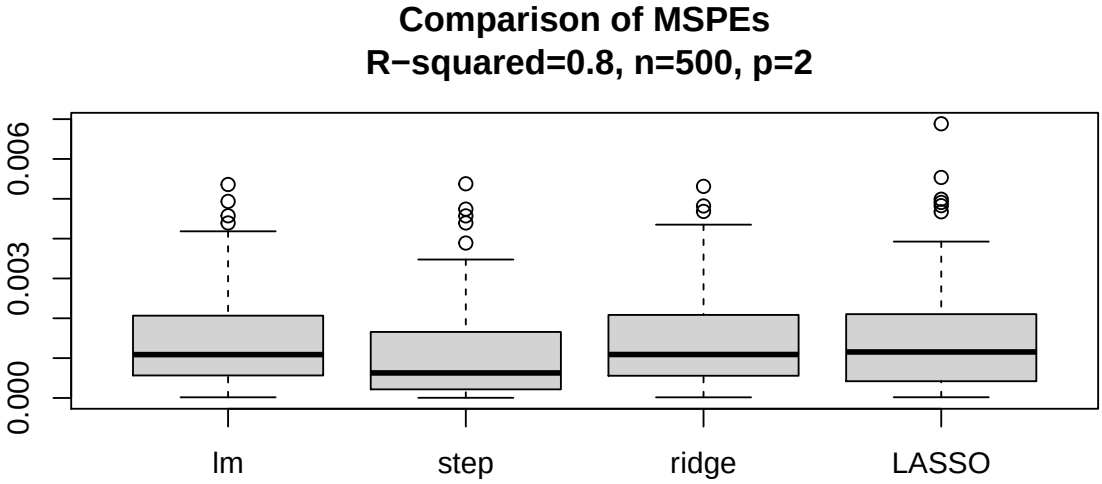
For 50/10/0.5, stepwise and Lasso look like the best model.

1.6 Question 6



For 500/100/3, LASSO is clearly the best model.

1.7 Question 7



For 500/2/0.5, basically everything is pretty similar. Stepwise MIGHT be a little better, but it's all very close.

2 Applications

2.1 Question 1

(a) The optimal λ value chosen by GCV is 0.2.
(b) The fitted coefficient values are as follows:

```
##       (Intercept) Solar.R Wind Temp   TWcp TWrat
## lm           -191  0.0638 9.56 2.89 -0.148  1.37
## Ridge        -162  0.0628 6.83 2.46 -0.109  1.65
```

Ignoring intercept, all but one ridge regression coefficient is smaller. Specifically, the coefficient on the temperature / wind speed ratio is larger for the optimal ridge regression model.

2.2 Question 2

Note that no seed is explicitly specified in this question. I used 4099183 for consistency with the Lecture 3 assignment.

(a) The chosen values of λ are: $\lambda_{min} = 1.35$, and $\lambda_{1se} = 10.42$. Rerunning the cv.glmnet fit many times without resetting seed, I get values between 0.0066–2.25 for λ_{min} and 5.96–13.78 for λ_{1se}.

(b) Fitted coefficient values for my λ values are:

```
## 2 x 6 sparse Matrix of class "dgCMatrix"
##           (Intercept) Solar.R Wind  Temp    TWcp TWrat
## min 1.35        -83.0  0.0484    . 1.280 -0.0075  2.39
## 1se 10.42       -31.5  .         . 0.742  .       1.68
```

```
## 2 x 6 sparse Matrix of class "dgCMatrix"
##            (Intercept) Solar.R Wind Temp   TWcp TWrat
## min 0.0066      -184.0 0.06350 8.88 2.79 -0.138  1.43
## 1se 5.96         -58.6 0.00824 .    1.01  .      2.18
```

```
## 2 x 6 sparse Matrix of class "dgCMatrix"
##           (Intercept) Solar.R   Wind  Temp TWcp TWrat
## min 2.25        -74.8  0.0404 -0.351 1.160    .  2.43
## 1se 13.78       -10.8  .       .     0.522    .  1.31
```

Regardless of the CV results, under the 1SE rule, more coefficients are set to zero, and the non-zero coefficients are mostly smaller.

(c) In the Lecture 5 assignment, hybrid stepwise selected Solar.R, Temp, and TWrat. Under the Min rule, all these variables remain active, along with at least one variable, depending on CV results. Under the 1SE rule, solar radiation may be excluded, but otherwise the active variables are the same as hybrid stepwise.

2.3 Question 3

(a) The MSPEs of ridge regression and both LASSO versions on each fold, as well as on the full dataset, are as follows.

```
##             1   2   3   4   5   6   7   8    9  10 Full
## Ridge     281 364 539 108 185 372 220 588 1040 725  442
## LASSO-Min 314 317 586 115 190 373 217 586 1050 651  439
## LASSO-1se 416 237 717 157 324 669 367 643 1130 378  503
```

(b)Boxplots of MSPEs and RMSPEs are given in Figures 1 and 2 respectively.

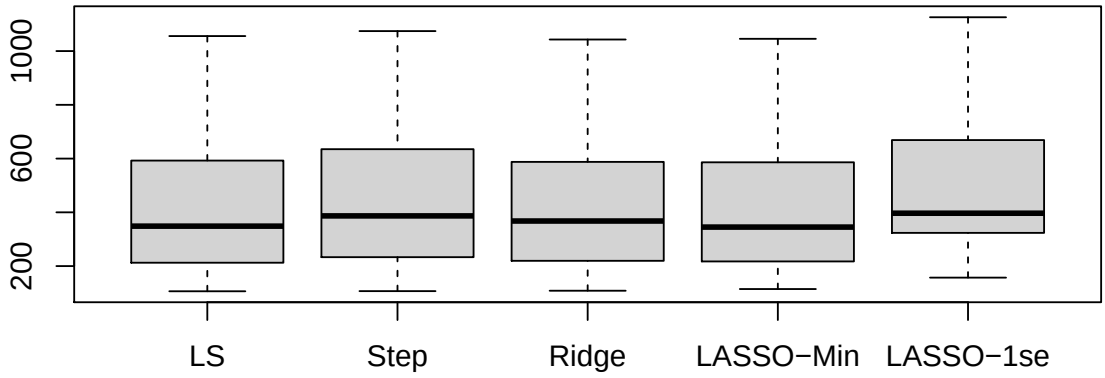


Figure 1: MSPE Boxplots

Based on the MSPEs in Figure 1, it is really hard to say that one model is clearly better than others. LS, ridge and LASSO-min look like the best models, just barely, with stepwise a bit worse, and LASSO-1se a bit worse still. That being said, all the models' boxplots have considerable overlap, and no model looks substantially better or worse.

(c) Based on the RMSPEs in Figure 2, LS and ridge seem to be performing best, with LS looking a bit better. Stepwise and LASSO-min are the next-best pair, with LASSO-min looking a bit better than stepwise. LASSO-1se is clearly the worst performing model.

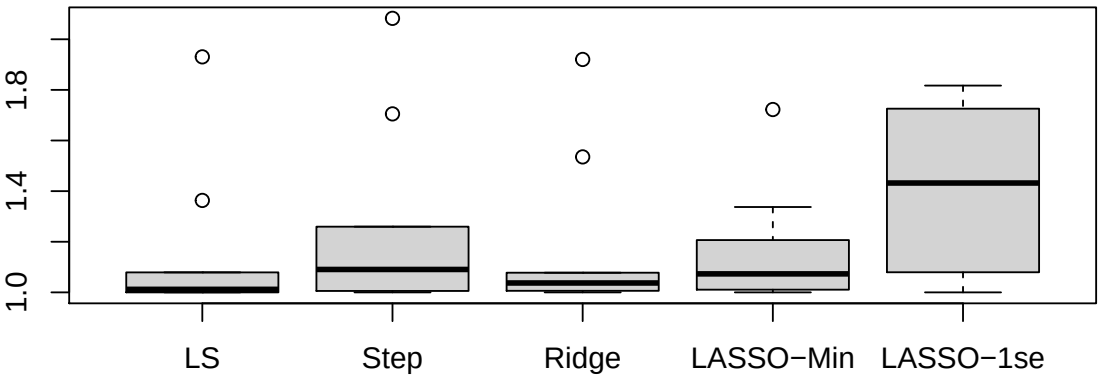


Figure 2: RMSPE Boxplots