ASM Practice

Ridge Regression

Maria Gkotsopoulou & Ricard Monge Calvo & Amalia Vradi

27/10/2019

Choosing the penalization parameter λ

In order to choose the best penalization parameter, we provide three functions that given a candidate list of parameters return the Mean Square Predictive Error (MSPE), performance measure of their associated ridge regression models.

Ridge regression lambda search

Our first function estimates the MSPE by computing the Mean Squared Error (MSE) of the actual target value y against the predicted target value $\hat{y}$ for a given validation set. Additionally, we plot the values of these MSPE against the $\log(1+\lambda)$ and the effective number of parameters df, or degrees of freedom.

Ridge regression lambda search with CV

Our second function performs a *k-fold* Cross Validation (CV) and averages the MSE of each fold to get the MSPE. As before, we plot the values of these MSPE against the $\log(1+\lambda)$ and the effective number of parameters df, or degrees of freedom.

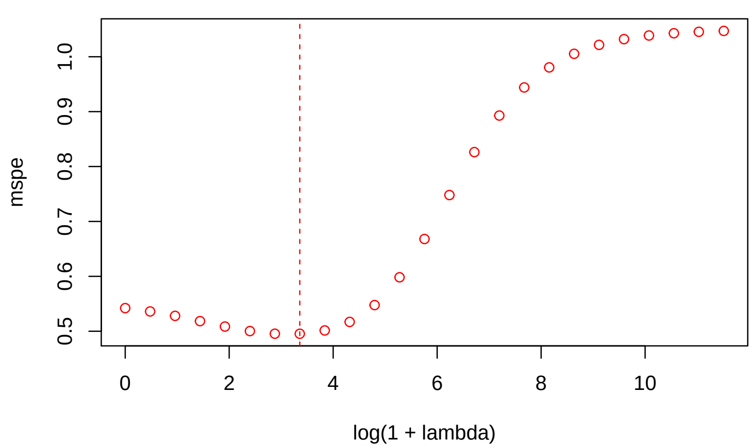
Prostate data application

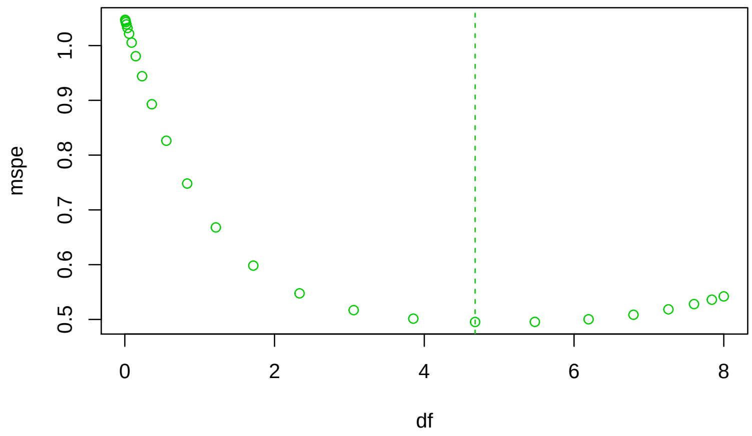
In addition to the previously mentioned functions, we provide a third one which computes the MSPE Leave One Out Cross Validation (LOOCV) and Generalized Cross Validation (GCV) estimates directly from the whole dataset.

To showcase these functions, we apply them to the *Prostate* dataset in the following three scenarios:

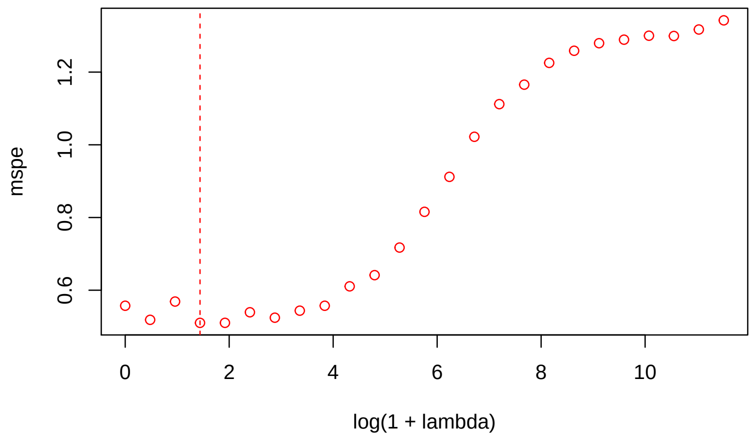
- Using a validation set of 30 instances (delimitted by the *train* variable).
- Using *5-fold* and *10-fold* Cross Validation.
- Using LOOCV (both as n-fold CV and the estimate) and GCV

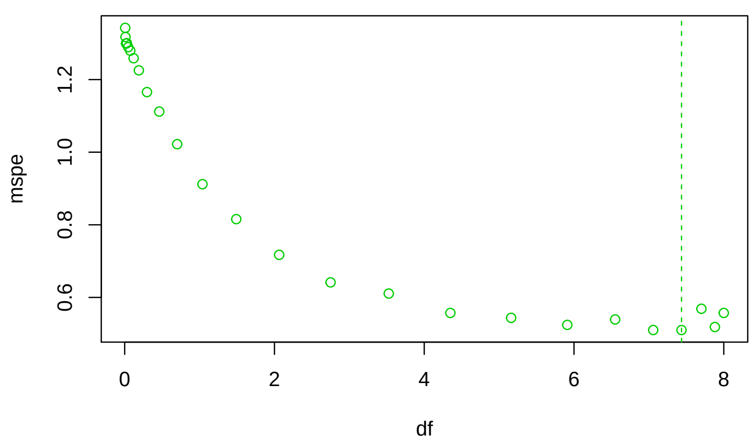
Validation set

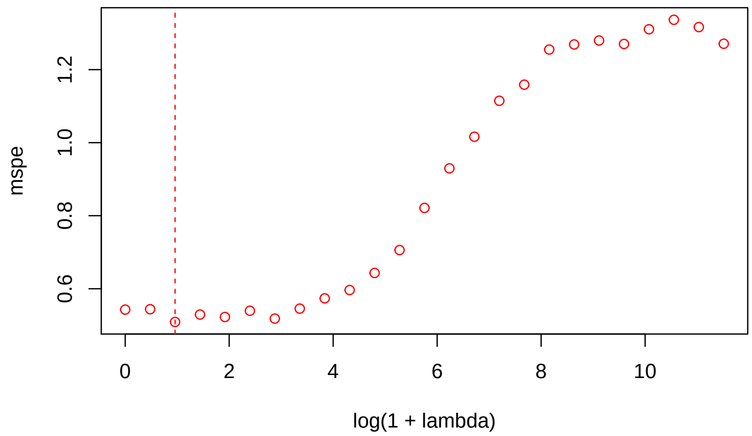


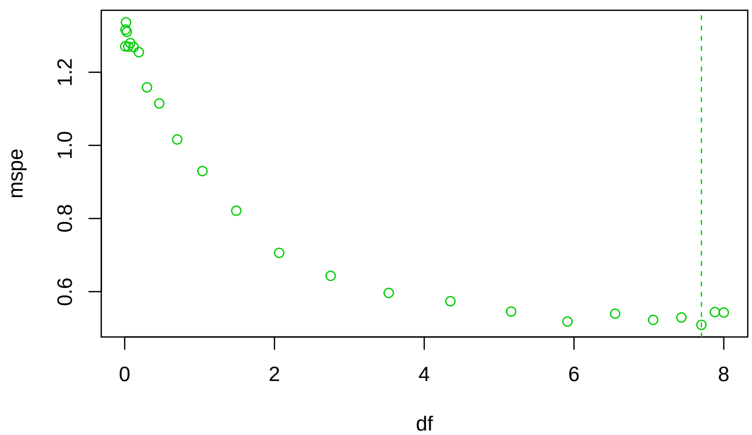


5-fold and 10-fold CV

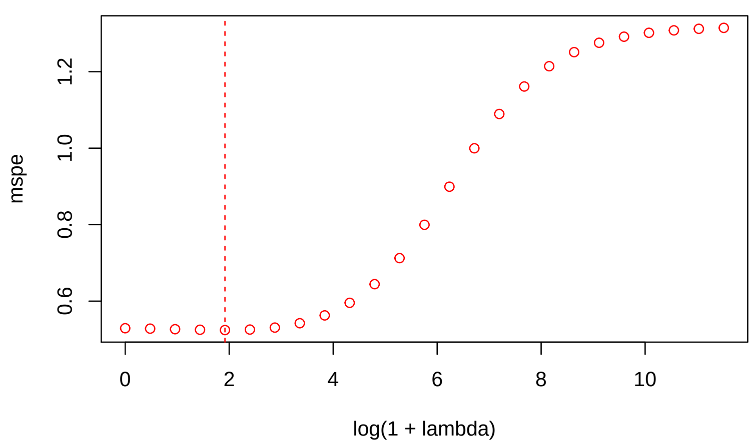


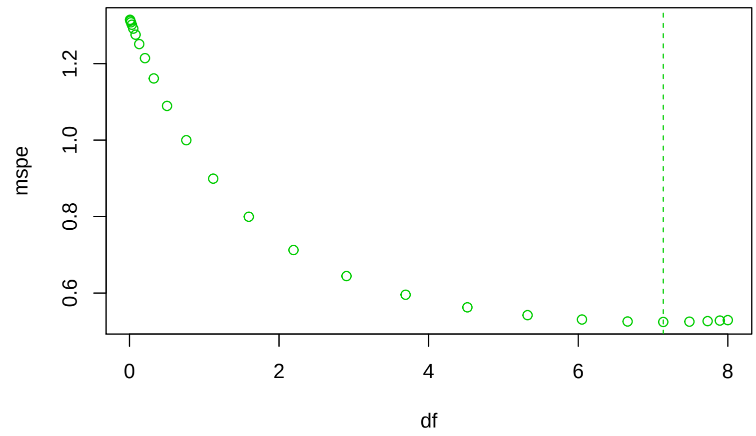


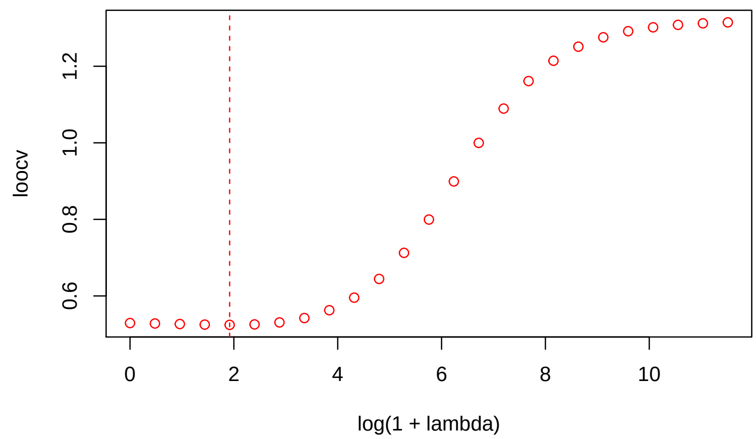


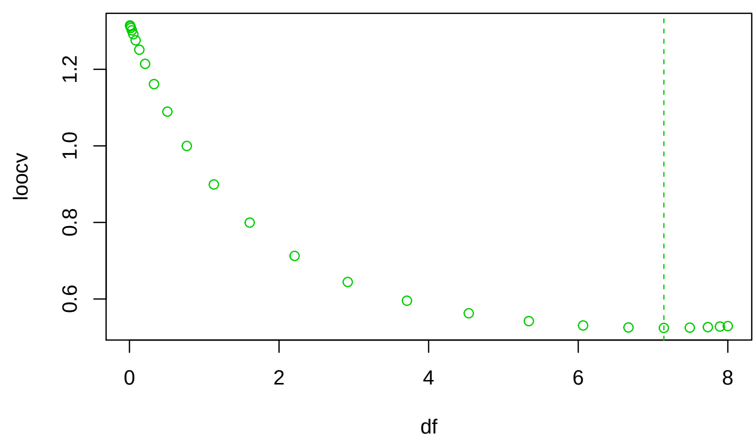


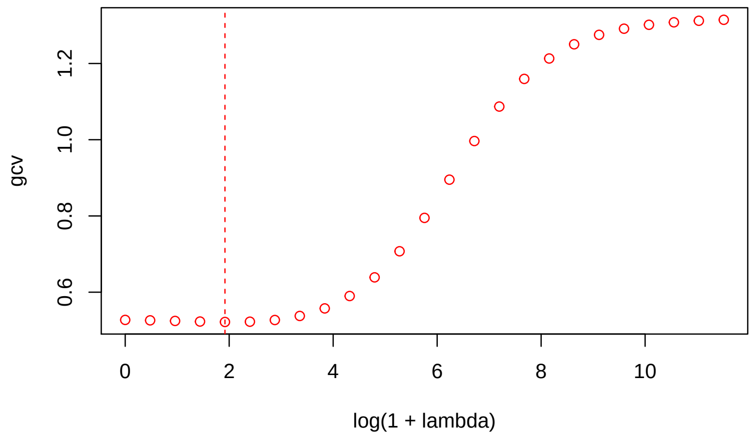
LOOCV (with n-fold CV) and GCV

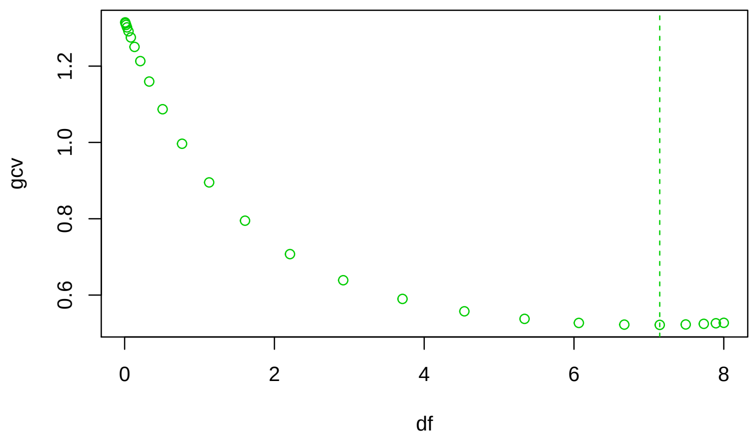



mspe
1.2
1.0
0.8
0.6
0
2
4
6
8
df


loocv
1.2
1.0
0.8
0.6
0
2
4
6
8
10
log(1 + lambda)


loocv
1.2
1.0
0.8
0.6
0
2
4
6
8
df

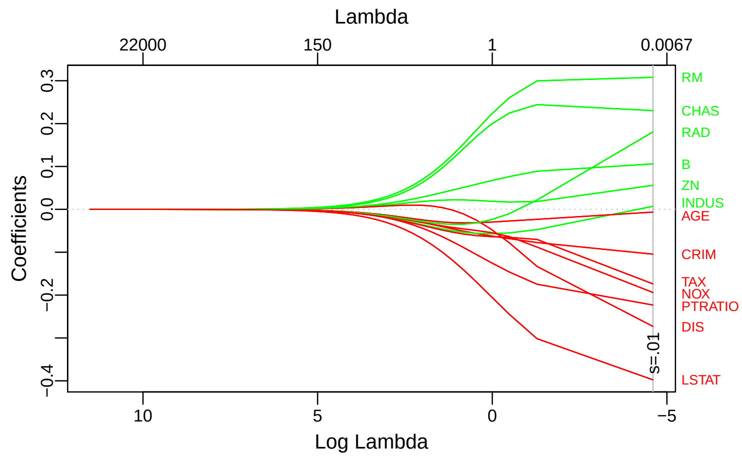




We observe that the results of MSPE seem more stable in the LOOCV and GCV compared to 5-fold and 10-fold CV, as one would expect given the small size of the dataset. Furthermore, the validation set results are also relatively stable, although with a different curve than for the LOOCV/GCV.

Ridge regression for the Boston Housing data

We start by scaling and splitting the Boston dataset to training and test using a 2/3 ratio. Since *CHAR* is a factor variable we do not include it in the *scale* function. First we need to tune the parameter λ. To do this we use 10 fold cross validation performed by *cv.glmnet*.



To select the best model, we now use 10x10-CV using the lambda that best minimised the error in cross-validation, which is 0.0100502.

So our final model has *Df*= 13 which is the number of non-zero coefficients and *%Dev*= 0.759315 is the percent deviance explained, which is quite good.

In terms of interpreting the coefficients, we observe that each additional room (*RM*) is associated with an increase in the house price, on average. This is quite straightforward, in principle, since it is to be expected that the larger the house, loosely speaking, the more expensive it will be. In addition, we see that an increase in *RAD* (index of accessibility to radial highways) is associated with an increse in *MEDV*. So basically, if we were to think of the town as a graph we would be capturing the connectivity degree of a specific suburb; so a remote node would have a lower value. Moreover, an increase in *CHAS* would mean that it will take the value of 1 is associate with an increase in *MEDV*, so ultimately if the Charles River passes through this suburb then this signals a higher house price, on average.

On the other hand, an increase in *LSTAT* (% lower status of the population) is associated with a decrease in the house price, on average. Most interestinglty though is that the increase in *PTRATIO* (pupil-teacher ratio by town) is associated with a decrease in the house price, on average. So in other words, the education offering of a town increases its value. Also, an increase in *DIS* (weighted distances to five Boston employment centres) is associated with a decrease in *MEDV*, on average. So, having to do a larger commute to work signals a lower house price. Another reasonable result is the fact that an increase in *NOX* (nitric oxides concentration) is associated with a decrease in *MEDV*, so air pollution is a detractor to house price.

Furthermore, we see that neither *AGE* (proportion of owner-occupied units built prior to 1940) nor *INDUS* (roportion of non-retail business acres per town) seem to have a considerable effect on *MEDV*.

Finally, we obtain the train and test error.

Table 1: Model Errors Summary

regression_method	train_MSE	test_MSE
ridge regression	0.245	0.309

The difference between train and test errors is not that large, even if the test set is relatively small, and thus subject to a great variance.