Power Method Examples

David Elsheimer

2/26/2020

Power Method

```r
y<- as.matrix(as.vector(read.csv("homework3_y.csv")))
X<-as.matrix(read.csv(gzfile("homework3_x.csv.gz")))
lambda<-10^seq(-5,5)
dim(y)
```

```
## [1] 58  1
```

```r
dim(X)
```

```
## [1]    58 12625
```

```r
kappa(X%*%t(X))
```

```
## [1] 156.492
```

```r
Betas<-dbelsheiST758::ridge_regression(y,X,lambda)
gcvlambda<-dbelsheiST758::gcv(y,X,lambda)
LOOCVlambda<-dbelsheiST758::leave_one_out(y,X,lambda)
gcvlambda
```

```
##            [,1]
##  [1,] 5367.639
##  [2,] 5336.307
##  [3,] 5334.678
##  [4,] 5334.359
##  [5,] 5334.332
##  [6,] 5334.329
##  [7,] 5334.328
##  [8,] 5334.326
##  [9,] 5334.297
## [10,] 5334.014
## [11,] 5331.187
```

```r
LOOCVlambda
```

```
##            [,1]
##  [1,] 4878.699
##  [2,] 4816.020
##  [3,] 4817.325
##  [4,] 4817.055
##  [5,] 4817.030
##  [6,] 4817.022
##  [7,] 4817.021
##  [8,] 4817.019
##  [9,] 4816.994
## [10,] 4816.742
## [11,] 4814.230
```

```r
plotdata<- as.data.frame(cbind(lambda,log10(lambda),gcvlambda,LOOCVlambda))
ggplot(plotdata, aes(x=log10(lambda),y=gcvlambda)) +geom_point()+
  xlab("log(lambda)") + ylab("GCV Criterion")+
  geom_vline(xintercept=log10(lambda)[gcvlambda == min(gcvlambda)],color='red')
```
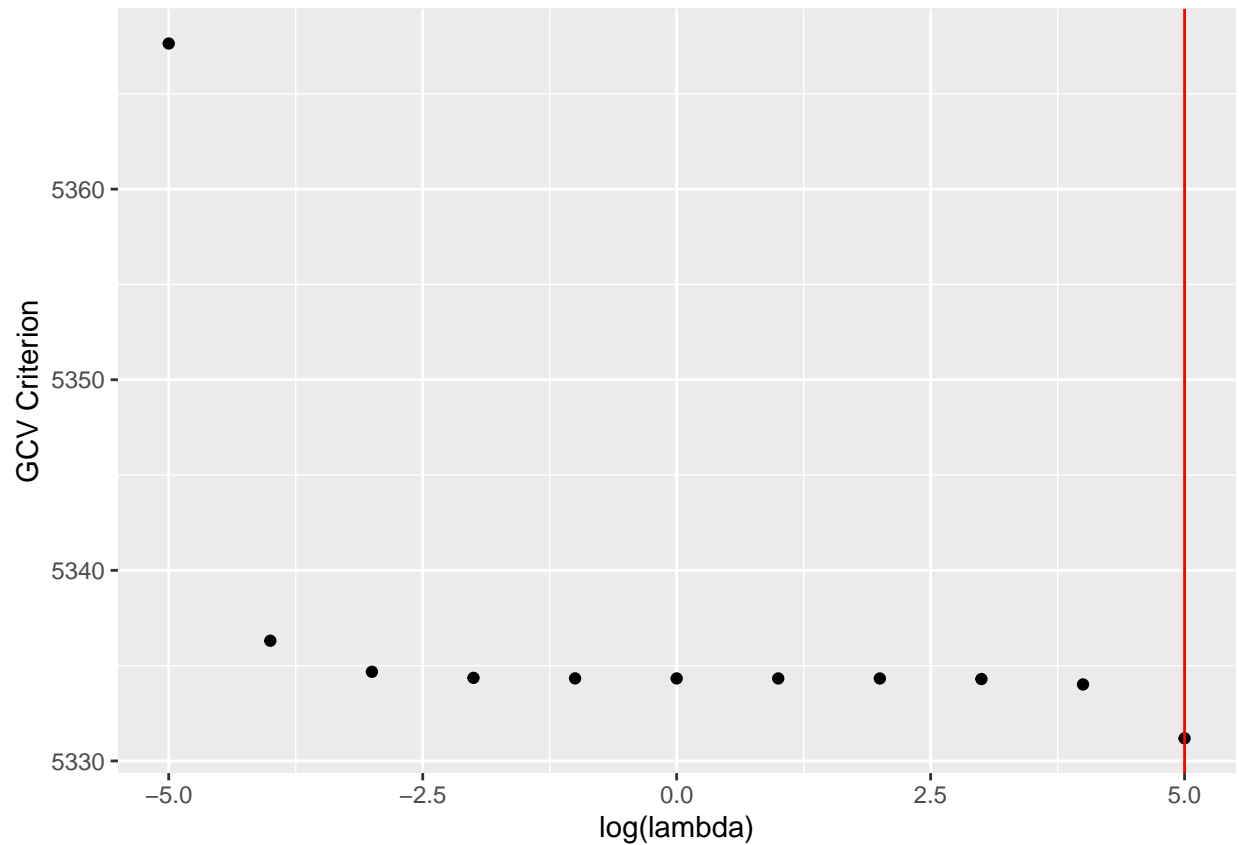


```r
ggplot(plotdata, aes(x=log10(lambda),y=LOOCVlambda)) + geom_point()+
  xlab("log(lambda)") + ylab("LOOCV Error")+
  geom_vline(xintercept=log10(lambda)[LOOCVlambda == min(LOOCVlambda)],color='red')
```
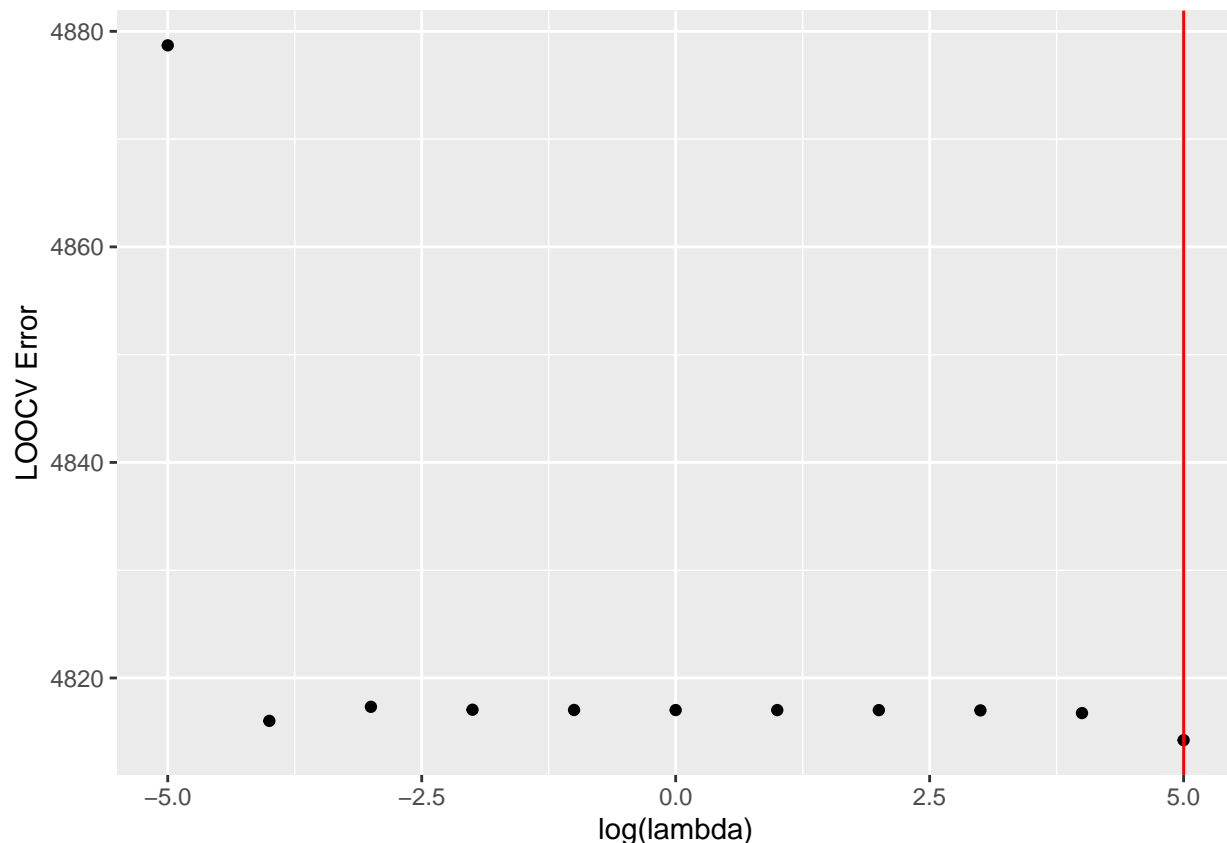
Based on the information from these two plots, GCV and LOOCV error are both minimized for $\lambda = 10^5$. In both cases the largest change occurs between $\lambda = 10^{-5}$ and $\lambda = 10^{-4}$. As LOOCV and GCV are not computationally intensive here, it is reasonable to choose the best tuning parameter based on the behavior of either in this example. To avoid problems with bias, it is best to choose a small tuning parameter. A large tuning parameter will increase bias, but due to the fact that $\mathbf{X}\mathbf{X}^\mathsf{T}$ has a relatively small condition number (around 156, found with `kappa()`), it will not be unstable for most perturbations. Thus the resulting matrix $\mathbf{X}\mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T} + \lambda\mathbf{I}_n)^{-1}$ will be nearly the identity matrix, resulting in a very small denominator in the GCV summation. Thus, the the introduction of large perturbations will be adding more bias but not greatly reducing variance. After $\lambda = 10^{-3}$, the reduction in GCV is little, so this is a reasonable choice for λ.