Ridge Regression: Prostate Cancer Data

Data

```
rm(list=ls())
# training and test data
require(ElemStatLearn)
train <- prostate[prostate$train,-10]
test <- prostate[!prostate$train,-10]
# n.ro of observations
n <- nrow(train)
m <- nrow(test)

# design matrix (without the intercept)
X<-as.matrix(
  scale(train[,-9]) # standardized
  )
X.star <- as.matrix(
  scale(test[,-9]) # standardized
  )
p = ncol(X)
# response
y = scale(train[,"lpsa"]) # standardized
y.star = scale(test[,"lpsa"]) # standardized
```

Exercize

The training and test data are

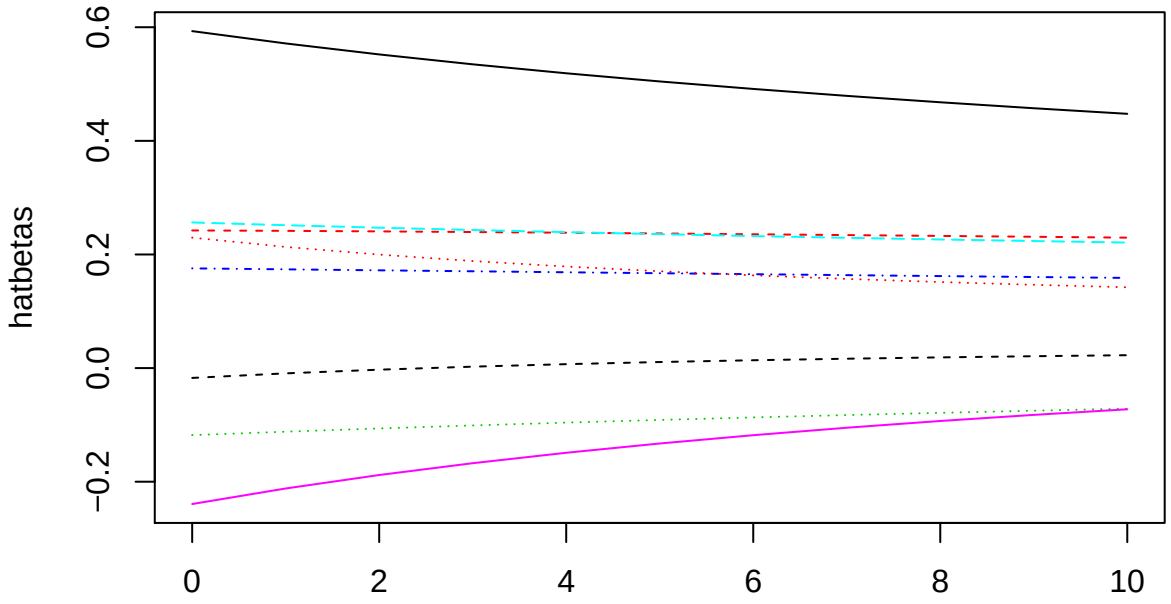
$$\underset{n\times 1}{\mathbf{y}}, \underset{n\times p}{\mathbf{X}}, \quad \underset{m\times 1}{\mathbf{y}^*}, \underset{m\times p}{\mathbf{X}^*}$$

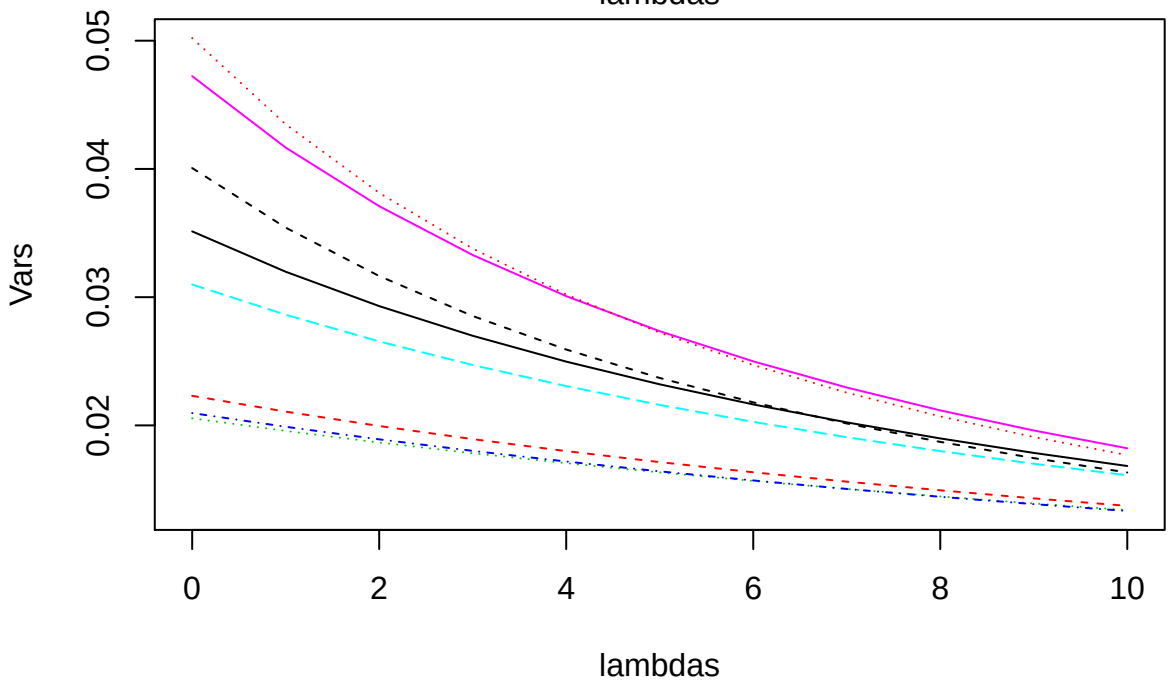
with $n = 67$, $m = 30$ and $p = 8$ for the Prostate Cancer data set, where the response and each predictor of the design matrix are standardized to have mean 0 and variance 1, separately for the training and test set.

The response variable is `lpsa`, and the predictor variables are `lcavol, lweight, age, lbph, svi, lcp, gleason, pgg45`.

1. Compute the rigde estimate
$$\hat{\beta}(\lambda) = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda\mathbf{I}_{p\times p})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$$
for $\lambda = 10$.
2. For $\lambda = 0, 1, 2, \ldots, 9, 10$, plot the solution path of the estimated coeffients $\hat{\beta}_1(\lambda), \ldots, \hat{\beta}_p(\lambda)$ and the corresponding variances $\mathbb{V}\text{ar}(\hat{\beta}_1(\lambda)), \ldots, \mathbb{V}\text{ar}(\hat{\beta}_p(\lambda))$, which are the elements in the diagonal of the variance/covariance matrix
$$\mathbb{V}\text{ar}(\hat{\beta}(\lambda)) = \sigma^2\mathbf{W}_\lambda(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{W}_\lambda^\mathsf{T}$$
where $\mathbf{W}_\lambda = [\mathbf{I}_{p\times p} + \lambda(\mathbf{Z}^\mathsf{T}\mathbf{X})^{-1}]^{-1}$ and is set to $\sigma^2 = 1$.
3. Compare the LOOCV estimate and the test MSE for the linear model and the ridge regression model with $\lambda = 5$.

```
##                [,1]
## lcavol   0.44761782
## lweight  0.22960005
## age     -0.07149012
## lbph     0.15871434
## svi      0.22098408
## lcp     -0.07256130
## gleason  0.02269095
## pgg45    0.14225969
```






```
## LOOCV ridge:  0.382395
```

```
## MSE ridge:  0.4565821
## LOOCV lm:  0.3867472
## MSE lm:  0.4747565
```