Quiz 2

Jing Leng (GSI: Jiahe)

December 7, 2014

a

Linear regresssion:

```
m1 <- lm(siri ~ ., data = fat.tr)
summary(m1)
```

```
##
## Call:
## lm(formula = siri ~ ., data = fat.tr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.328  -2.944  -0.105   2.909   9.665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.8209    17.9830   -1.10   0.2716
## age           0.0672     0.0341    1.97   0.0501 .
## weight       -0.0956     0.0556   -1.72   0.0872 .
## height       -0.0446     0.1123   -0.40   0.6918
## adipos       -0.0491     0.3164   -0.16   0.8767
## neck         -0.4380     0.2485   -1.76   0.0794 .
## chest        -0.0824     0.1094   -0.75   0.4522
## abdom         1.0302     0.0978   10.53   <2e-16 ***
## hip          -0.2041     0.1557   -1.31   0.1914
## thigh         0.2536     0.1519    1.67   0.0964 .
## knee          0.0297     0.2609    0.11   0.9094
## ankle         0.1572     0.2268    0.69   0.4889
## biceps        0.1897     0.1802    1.05   0.2939
## forearm       0.4677     0.2038    2.29   0.0228 *
## wrist        -1.7432     0.5601   -3.11   0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.32 on 212 degrees of freedom
## Multiple R-squared:  0.759,  Adjusted R-squared:  0.743
## F-statistic: 47.7 on 14 and 212 DF,  p-value: <2e-16
```

```
m1.rmse.tr <- rmse(m1$fitted.values, fat.tr$siri); m1.rmse.tr
```

```
## [1] 4.179
```

```
ypred <- predict(m1, fat.ts[-1])
m1.rmse.ts <- rmse(ypred, fat.ts$siri); m1.rmse.ts
```

```
## [1] 4.396
```

b

```
m2 <- step(m1)
```

```
## Start:  AIC=679.2
## siri ~ age + weight + height + adipos + neck + chest + abdom +
##     hip + thigh + knee + ankle + biceps + forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## - knee     1         0 3964 677
## - adipos   1         0 3964 677
## - height   1         3 3967 677
## - ankle    1         9 3973 678
## - chest    1        11 3974 678
## - biceps   1        21 3984 678
## - hip      1        32 3996 679
## <none>                 3964 679
## - thigh    1        52 4016 680
## - weight   1        55 4019 680
## - neck     1        58 4022 681
## - age      1        73 4036 681
## - forearm  1        98 4062 683
## - wrist    1       181 4145 687
## - abdom    1      2074 6038 773
##
## Step:  AIC=677.2
## siri ~ age + weight + height + adipos + neck + chest + abdom +
##     hip + thigh + ankle + biceps + forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## - adipos   1         1 3965 675
## - height   1         3 3967 675
## - ankle    1        10 3974 676
## - chest    1        11 3975 676
## - biceps   1        21 3984 676
## - hip      1        32 3996 677
## <none>                 3964 677
## - weight   1        58 4021 679
## - neck     1        60 4024 679
## - thigh    1        61 4025 679
## - age      1        81 4045 680
## - forearm  1       100 4064 681
## - wrist    1       181 4145 685
## - abdom    1      2074 6038 771
##
## Step:  AIC=675.3
## siri ~ age + weight + height + neck + chest + abdom + hip + thigh +
##     ankle + biceps + forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## - height   1         2 3967 673
## - ankle    1         9 3974 674
## - chest    1        14 3979 674
```

```
## - biceps   1        20 3984 674
## - hip      1        35 3999 675
## <none>                 3965 675
## - weight   1        58 4022 677
## - thigh    1        60 4025 677
## - neck     1        65 4029 677
## - age      1        83 4047 678
## - forearm  1       100 4064 679
## - wrist    1       181 4146 683
## - abdom    1      2332 6296 778
##
## Step:  AIC=673.4
## siri ~ age + weight + neck + chest + abdom + hip + thigh + ankle +
##     biceps + forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## - ankle    1        10 3977 672
## - chest    1        12 3979 672
## - biceps   1        21 3987 673
## - hip      1        32 3999 673
## <none>                 3967 673
## - neck     1        64 4031 675
## - thigh    1        67 4034 675
## - age      1        87 4054 676
## - weight   1        91 4058 677
## - forearm  1       100 4067 677
## - wrist    1       188 4155 682
## - abdom    1      2374 6341 778
##
## Step:  AIC=672
## siri ~ age + weight + neck + chest + abdom + hip + thigh + biceps +
##     forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## - chest    1        13 3990 671
## - biceps   1        19 3996 671
## - hip      1        33 4010 672
## <none>                 3977 672
## - thigh    1        69 4046 674
## - neck     1        71 4048 674
## - weight   1        82 4059 675
## - age      1        84 4061 675
## - forearm  1       100 4076 676
## - wrist    1       178 4155 680
## - abdom    1      2368 6345 776
##
## Step:  AIC=670.7
## siri ~ age + weight + neck + abdom + hip + thigh + biceps + forearm +
##     wrist
##
##           Df Sum of Sq  RSS AIC
## - biceps   1        16 4006 670
## - hip      1        28 4018 670
## <none>                 3990 671
```

```
## - neck     1        76 4066 673
## - thigh    1        81 4071 673
## - age      1        81 4071 673
## - forearm  1        93 4082 674
## - weight   1       121 4111 675
## - wrist    1       173 4163 678
## - abdom    1      2963 6953 795
##
## Step:  AIC=669.6
## siri ~ age + weight + neck + abdom + hip + thigh + forearm +
##     wrist
##
##           Df Sum of Sq  RSS AIC
## - hip      1        32 4038 669
## <none>                 4006 670
## - neck     1        71 4076 672
## - age      1        88 4094 673
## - weight   1       108 4114 674
## - thigh    1       108 4114 674
## - forearm  1       131 4137 675
## - wrist    1       172 4178 677
## - abdom    1      2947 6953 793
##
## Step:  AIC=669.4
## siri ~ age + weight + neck + abdom + thigh + forearm + wrist
##
##           Df Sum of Sq  RSS AIC
## <none>                 4038 669
## - neck     1        54 4092 670
## - thigh    1        77 4115 672
## - age      1        93 4131 673
## - forearm  1       150 4188 676
## - wrist    1       174 4212 677
## - weight   1       240 4278 681
## - abdom    1      3006 7044 794
```

```
summary(m2)
```

```
##
## Call:
## lm(formula = siri ~ age + weight + neck + abdom + thigh + forearm +
##     wrist, data = fat.tr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.172  -3.125  -0.264   3.089   9.315
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33.7921     9.4305   -3.58  0.00042 ***
## age           0.0718     0.0320    2.24  0.02587 *
## weight       -0.1279     0.0355   -3.61  0.00039 ***
## neck         -0.3962     0.2312   -1.71  0.08798 .
## abdom         0.9487     0.0743   12.77  < 2e-16 ***
```

```
## thigh         0.2422     0.1183    2.05  0.04178 *
## forearm       0.5398     0.1891    2.85  0.00472 **
## wrist        -1.6373     0.5337   -3.07  0.00243 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.29 on 219 degrees of freedom
## Multiple R-squared:  0.755,  Adjusted R-squared:  0.747
## F-statistic: 96.2 on 7 and 219 DF,  p-value: <2e-16
```

```
m2.rmse.tr <- rmse(m2$fitted.values, fat.tr$siri); m2.rmse.tr
```

```
## [1] 4.218
```

```
ypred <- predict(m2, fat.ts[-1])
m2.rmse.ts <- rmse(ypred, fat.ts$siri); m2.rmse.ts
```

```
## [1] 4.342
```

c

```
library(pls)
```

```
##
## Attaching package: 'pls'
##
## The following object is masked from 'package:stats':
##
##     loadings
```

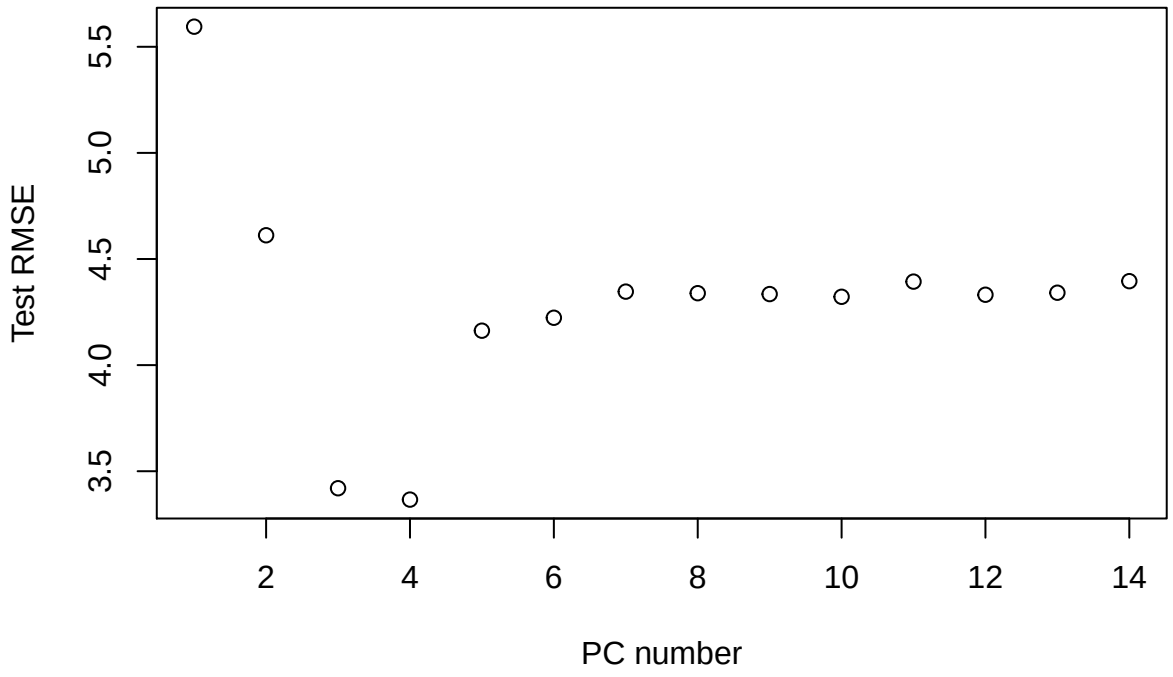
```
m3 <- pcr(siri ~ ., data=fat.tr, ncomp=14)
rmsemeat <- NULL
for (k in 1:14) {
  pv <- predict(m3, newdata=fat.ts, ncomp=k)
  rmsemeat[k] <- rmse(pv, fat.ts$siri)
}
min(rmsemeat)
```

```
## [1] 3.366
```

```
which.min(rmsemeat)
```

```
## [1] 4
```

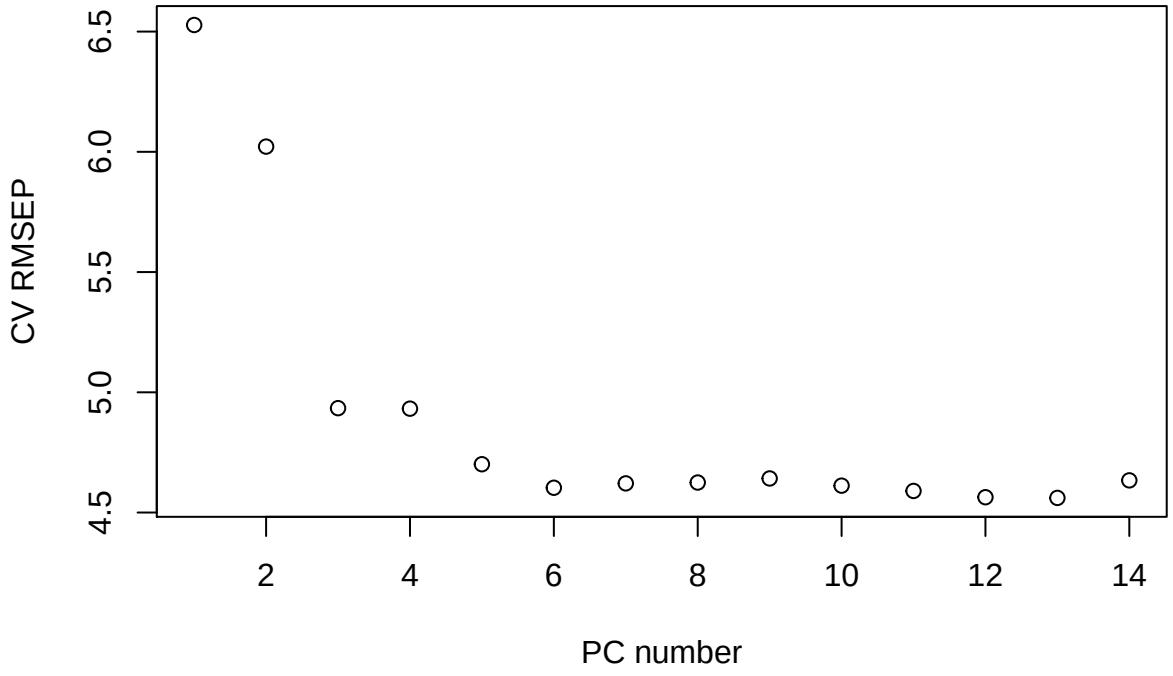
```
plot(rmsemeat, xlab="PC number", ylab="Test RMSE")
```




```
m4 <- pcr(siri ~ ., data=fat.tr, ncomp=14,
             validation="CV", segments = 10)
rmseCV <- RMSEP(m4, estimate="CV", intercept=F)
which.min(rmseCV$val)
```

```
## [1] 13
```

```
m4.rmse.tr <- min(rmseCV$val)
plot(rmseCV$val, xlab="PC number", ylab="CV RMSEP")
```



```r
yfit <- predict(m4, newdata=fat.ts, ncomp=which.min(rmseCV$val))
m4.rmse.ts <- rmse(yfit, fat.ts$siri); m4.rmse.ts
```

```
## [1] 4.341
```

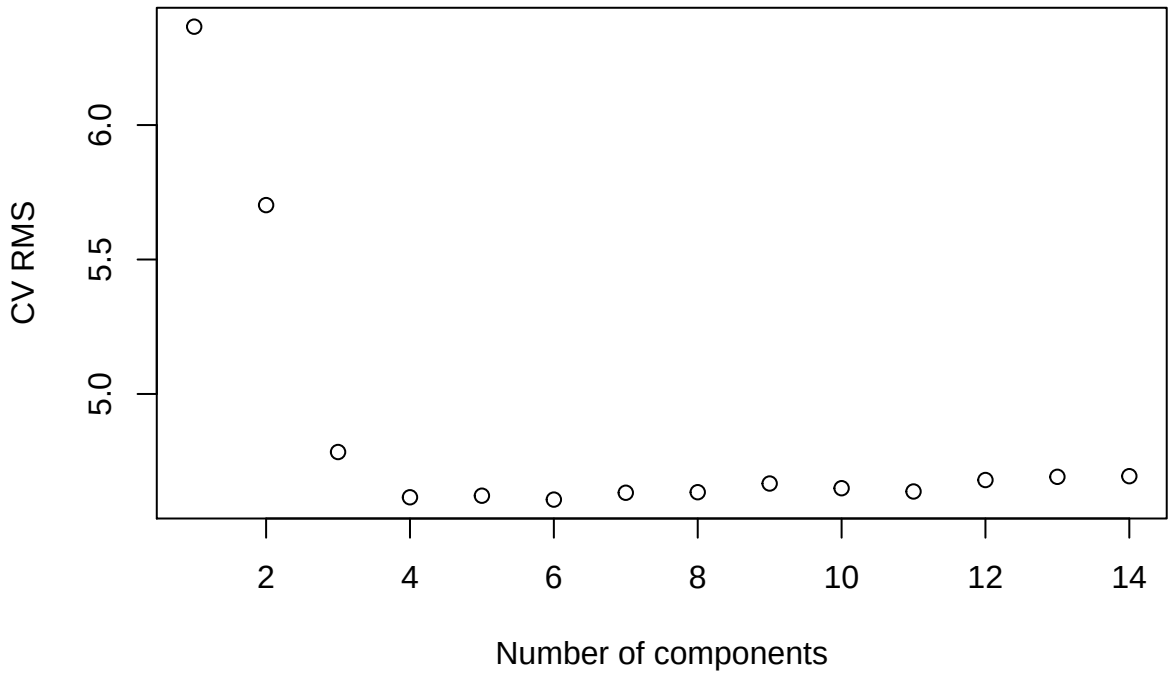
d

```r
m5 <- plsr(siri ~., data = fat.tr, ncomp = 14, validation = "CV")
summary(m5)
```

```
## Data:    X dimension: 227 14
##  Y dimension: 227 1
## Fit method: kernelpls
## Number of components considered: 14
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV           8.551    6.366    5.702    4.784    4.616    4.622    4.607
## adjCV        8.551    6.359    5.694    4.773    4.599    4.597    4.586
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV       4.633    4.635    4.667      4.65     4.637     4.680     4.692
## adjCV    4.609    4.610    4.636      4.62     4.610     4.648     4.659
##        14 comps
## CV        4.694
## adjCV     4.661
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X       82.92    95.00    97.16    97.75    98.11    98.75     99.1
## siri    46.32    57.76    71.05    73.96    74.88    75.15     75.4
##       8 comps  9 comps  10 comps  11 comps  12 comps  13 comps  14 comps
## X       99.32    99.44     99.54     99.67     99.81     99.89    100.00
## siri    75.60    75.84     75.90     75.91     75.91     75.91     75.91
```

```r
pls_rmsCV <- RMSEP(m5, estimate="CV", intercept=F)

plot(pls_rmsCV$val, xlab="Number of components",
     ylab="CV RMS")
```




```
m5.rmse.tr <- min(pls_rmsCV$val); m5.rmse.tr
```

```
## [1] 4.607
```

```
which.min(pls_rmsCV$val)
```

```
## [1] 6
```

```
ytpred <- predict(m5, fat.ts, ncomp=which.min(pls_rmsCV$val))
m5.rmse.ts <- rmse(ytpred, fat.ts$siri); m5.rmse.ts
```

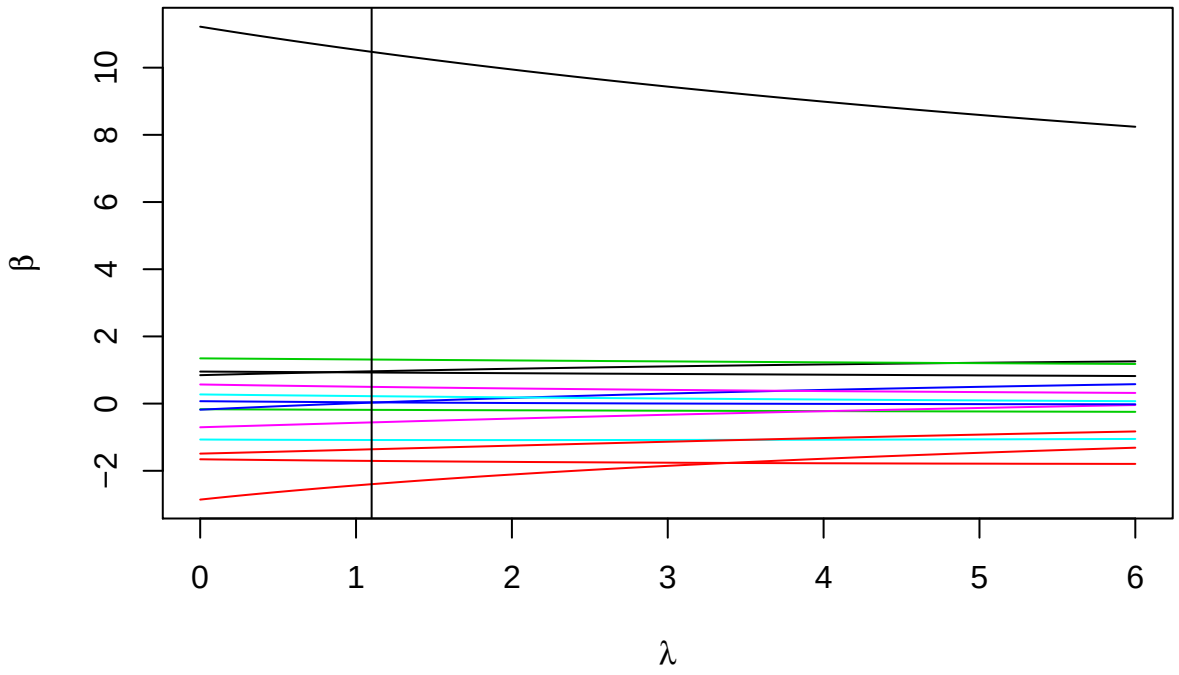
```
## [1] 4.332
```

e

```
library(MASS)
m6 = lm.ridge(siri ~ ., lambda=seq(0, 6, 0.1),
                  data = fat.tr)
select(m6)
```

```
## modified HKB estimator is 1.552
## modified L-W estimator is 4.078
## smallest value of GCV  at 1.1
```

```
matplot(m6$lambda, t(m6$coef), type="l", lty=1,
        xlab=expression(lambda), ylab=expression(hat(beta)))
abline(v=1.1)
```




```
which.min(m6$GCV)
```

```
## 1.1
##  12
```

```
yfit <- m6$ym + scale(fat.tr[,-1], center=m6$xm,
                          scale=m6$scales ) %*% m6$coef[, 12]
m6.rmse.tr <- rmse(yfit, fat.tr$siri); m6.rmse.tr
```

```
## [1] 4.184
```

```
ypred <- m6$ym + scale(fat.ts[,-1], center=m6$xm,
                          scale = m6$scales ) %*% m6$coef[,12]
m6.rmse.ts <- rmse(ypred, fat.ts$siri); m6.rmse.ts
```

```
## [1] 4.282
```

Sample	LS	AIC	PCR	PRS	Ridge
Training	4.1787	4.2177	4.5609	4.6072	4.1839
Test	4.3956	4.3425	4.3411	4.3323	4.2816