

HW#10, Nan Deng

1)

a)

```
library(faraway)
data(fat)
index <- seq(10, 250, by=10)
train <- fat[-index, -c(1,3,8)]
test <- fat[index, -c(1,3,8)]
rmse <- function(x,y) sqrt(mean((x-y)^2))
fat_fit <- lm(siri ~ ., data=train)
#summary(fat_fit)
rmse(fat_fit$fitted.values, train$siri)

## [1] 4.178651

rmse(predict(fat_fit, test), test$siri)

## [1] 4.395559
```

b)

```
fat_aic <- step(fat_fit,k=2)

## Start: AIC=679.21
## siri ~ age + weight + height + adipos + neck + chest + abdom +
## hip + thigh + knee + ankle + biceps + forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## - knee      1      0.24 3963.9 677.23
## - adipos     1      0.45 3964.1 677.24
## - height     1      2.95 3966.6 677.38
## - ankle      1      8.99 3972.7 677.73
## - chest      1     10.61 3974.3 677.82
## - biceps     1     20.70 3984.4 678.40
## - hip        1     32.11 3995.8 679.05
## <none>                3963.7 679.21
## - thigh      1     52.13 4015.8 680.18
## - weight     1     55.21 4018.9 680.35
## - neck       1     58.10 4021.8 680.52
## - age        1     72.57 4036.2 681.33
## - forearm    1     98.42 4062.1 682.78
## - wrist      1    181.11 4144.8 687.36
## - abdom      1   2074.23 6037.9 772.76
##
## Step: AIC=677.23
## siri ~ age + weight + height + adipos + neck + chest + abdom +
## hip + thigh + ankle + biceps + forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## - adipos     1      0.63 3964.6 675.26
## - height     1      3.00 3966.9 675.40
## - ankle      1      9.81 3973.7 675.79
## - chest      1     10.60 3974.5 675.84
## - biceps     1     20.53 3984.4 676.40
```

```

## - hip      1      31.87 3995.8 677.05
## <none>                3963.9 677.23
## - weight   1      57.56 4021.5 678.50
## - neck     1      59.81 4023.7 678.63
## - thigh    1      60.61 4024.5 678.67
## - age      1      81.08 4045.0 679.82
## - forearm  1      99.73 4063.6 680.87
## - wrist    1     181.04 4145.0 685.37
## - abdom    1     2074.02 6037.9 770.76
##
## Step:  AIC=675.26
## siri ~ age + weight + height + neck + chest + abdom + hip + thigh +
##      ankle + biceps + forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## - height   1         2.37 3966.9 673.40
## - ankle    1         9.44 3974.0 673.80
## - chest     1        14.11 3978.7 674.07
## - biceps    1        19.95 3984.5 674.40
## - hip       1        34.56 3999.1 675.24
## <none>                3964.6 675.26
## - weight   1        57.80 4022.3 676.55
## - thigh    1        59.98 4024.5 676.67
## - neck     1        64.73 4029.3 676.94
## - age      1        82.66 4047.2 677.95
## - forearm  1        99.78 4064.3 678.91
## - wrist    1       181.22 4145.8 683.41
## - abdom    1     2331.59 6296.1 778.26
##
## Step:  AIC=673.4
## siri ~ age + weight + neck + chest + abdom + hip + thigh + ankle +
##      biceps + forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## - ankle    1         9.80 3976.7 671.96
## - chest     1        12.32 3979.2 672.10
## - biceps    1        20.57 3987.5 672.57
## - hip       1        32.19 3999.1 673.24
## <none>                3966.9 673.40
## - neck     1        63.84 4030.8 675.02
## - thigh    1        66.75 4033.7 675.19
## - age      1        86.87 4053.8 676.32
## - weight   1        91.03 4058.0 676.55
## - forearm  1       100.16 4067.1 677.06
## - wrist    1       188.13 4155.1 681.92
## - abdom    1     2373.71 6340.6 777.86
##
## Step:  AIC=671.96
## siri ~ age + weight + neck + chest + abdom + hip + thigh + biceps +
##      forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## - chest     1        12.98 3989.7 670.70
## - biceps    1        19.28 3996.0 671.06
## - hip       1        32.95 4009.7 671.83

```

```

## <none>                3976.7 671.96
## - thigh      1      69.41 4046.1 673.89
## - neck       1      71.43 4048.2 674.00
## - weight     1      82.05 4058.8 674.60
## - age        1      84.39 4061.1 674.73
## - forearm    1      99.60 4076.3 675.58
## - wrist      1     178.38 4155.1 679.92
## - abdom      1    2367.81 6344.5 776.00
##
## Step:  AIC=670.7
## siri ~ age + weight + neck + abdom + hip + thigh + biceps + forearm +
##      wrist
##
##           Df Sum of Sq    RSS    AIC
## - biceps    1      16.22 4005.9 669.62
## - hip       1      28.00 4017.7 670.29
## <none>                3989.7 670.70
## - neck      1      76.13 4065.8 672.99
## - thigh     1      80.91 4070.6 673.26
## - age       1      80.92 4070.6 673.26
## - forearm   1      92.71 4082.4 673.91
## - weight    1     121.11 4110.8 675.49
## - wrist     1     173.11 4162.8 678.34
## - abdom     1    2963.15 6952.8 794.78
##
## Step:  AIC=669.62
## siri ~ age + weight + neck + abdom + hip + thigh + forearm +
##      wrist
##
##           Df Sum of Sq    RSS    AIC
## - hip       1      32.15 4038.1 669.44
## <none>                4005.9 669.62
## - neck      1      70.53 4076.5 671.58
## - age       1      88.36 4094.3 672.57
## - weight    1     107.79 4113.7 673.65
## - thigh     1     108.44 4114.4 673.68
## - forearm   1     130.86 4136.8 674.92
## - wrist     1     172.31 4178.2 677.18
## - abdom     1    2946.95 6952.9 792.79
##
## Step:  AIC=669.44
## siri ~ age + weight + neck + abdom + thigh + forearm + wrist
##
##           Df Sum of Sq    RSS    AIC
## <none>                4038.1 669.44
## - neck      1      54.16 4092.2 670.46
## - thigh     1      77.32 4115.4 671.74
## - age       1      92.80 4130.9 672.59
## - forearm   1     150.29 4188.4 675.73
## - wrist     1     173.55 4211.6 676.99
## - weight    1     239.75 4277.8 680.53
## - abdom     1    3006.10 7044.2 793.75

# summary(fat_aic)
rmse(fat_aic$fitted.values, train$siri)

```

```
## [1] 4.217687

rmse(predict(fat_aic, test), test$siri)

## [1] 4.342456
```

```
c)
library(pls)

##
## Attaching package: 'pls'

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

set.seed(123)
fat_pcr <- pcr(siri ~ ., data=train, comp=14, validation="CV", segments=10)
rmseCV <- RMSEP(fat_pcr, estimate = "CV", intercept = F)
which.min(rmseCV$val)

## [1] 13

rmse(fat_pcr$fitted.values[, , 13], train$siri)

## [1] 4.260173

yfit <- predict(fat_pcr, newdata = test, ncomp = 13)
rmse(yfit, test$siri)

## [1] 4.341128
```

```
d)
set.seed(123)
fat_pls <- plsr(siri ~ ., data=train, comp=14, validation="CV", segments=10)
rmseCV <- RMSEP(fat_pls, estimate = "CV", intercept = F)
which.min(rmseCV$val)

## [1] 4

rmse(fat_pls$fitted.values[, , 4], train$siri)

## [1] 4.344006

yfit <- predict(fat_pls, newdata = test, ncomp = 4)
rmse(yfit, test$siri)

## [1] 4.392838
```

```
e)
library(MASS)
fat_ridge <- lm.ridge(siri ~ ., lambda=seq(0, 1.2, 0.1), data=train)
select(fat_ridge)

## modified HKB estimator is 1.552127
## modified L-W estimator is 4.078233
## smallest value of GCV at 1.1

which.min(fat_ridge$GCV)
```

```
## 1.1
## 12

yfit <- fat_ridge$ym + scale(train[, -1], center = fat_ridge$xm, scale = fat_ridge$scales
) %*% fat_ridge$coef[, 12]
rmse(yfit, train$siri)

## [1] 4.183926

ypred <- fat_ridge$ym + scale(test[, -1], center = fat_ridge$xm, scale = fat_ridge$scales
) %*% fat_ridge$coef[, 12]
rmse(ypred, test$siri)

## [1] 4.281613
```

Sample	LS	AIC	PCR	PLS	Ridge
Training	4.178651	4.217687	4.260173	4.344006	4.183926
Test	4.395559	4.342456	4.341128	4.392838	4.281613

According to the calculation of RMSEs, Least Square Model and Partial Least Square Model have poorest performances. In comparison, LS model with AIC selection variables and Principle Component Model perform relatively better, while Ridge regression has the best performance among all models.