```r
library(lmridge)
library(MASS)
```

```r
# 输入数据
x1 <- c(82.9,88.0,99.9,105.3,117.7,131.0,148.2,161.8,174.2,184.7)
x2 <- c(92.0,93.0,96.0,94.0,100.0,101.0,105.0,112.0,112.0,112.0)
x3 <- c(17.1,21.3,25.1,29.0,34.0,40.0,44.0,49.0,51.0,53.0)
x4 <- c(94.0,96.0,97.0,97.0,100.0,101.0,104.0,109.0,111.0,111.0)
y <- c(8.4,9.6,10.4,11.4,12.2,14.2,15.8,17.9,19.6,20.8)
# 标准化设计阵与响应变量
n <- 10
X <- cbind(x1,x2,x3,x4)
Xs <- scale(X)/sqrt(n-1)
Ys <- scale(y)/sqrt(n-1)
```

```r
# 计算均值与样本标准方差
m1 <- mean(x1); m2 <- mean(x2); m3 <- mean(x3); m4 <- mean(x4); my <- mean(y)
s1 <- sqrt(sum(x1^2)-n*m1^2); s2 <- sqrt(sum(x2^2)-n*m2^2); s3 <- sqrt(sum(x3^2)-n*m3^2)
s4 <- sqrt(sum(x4^2)-n*m4^2); sy <- sqrt(sum(y^2)-n*my^2)
c(m1,m2,m3,m4,my)
```

```
## [1] 129.37 101.70  36.35 102.00  14.03
```

```r
c(s1,s2,s3,s4,sy)
```

```
## [1] 109.24761  24.37417  38.81862  19.23538  13.02924
```

```r
# 计算相关系数矩阵的特征值
R <- cor(X)
eigen(R)$values
```

```
## [1] 3.943813553 0.039962805 0.012586520 0.003637121
```

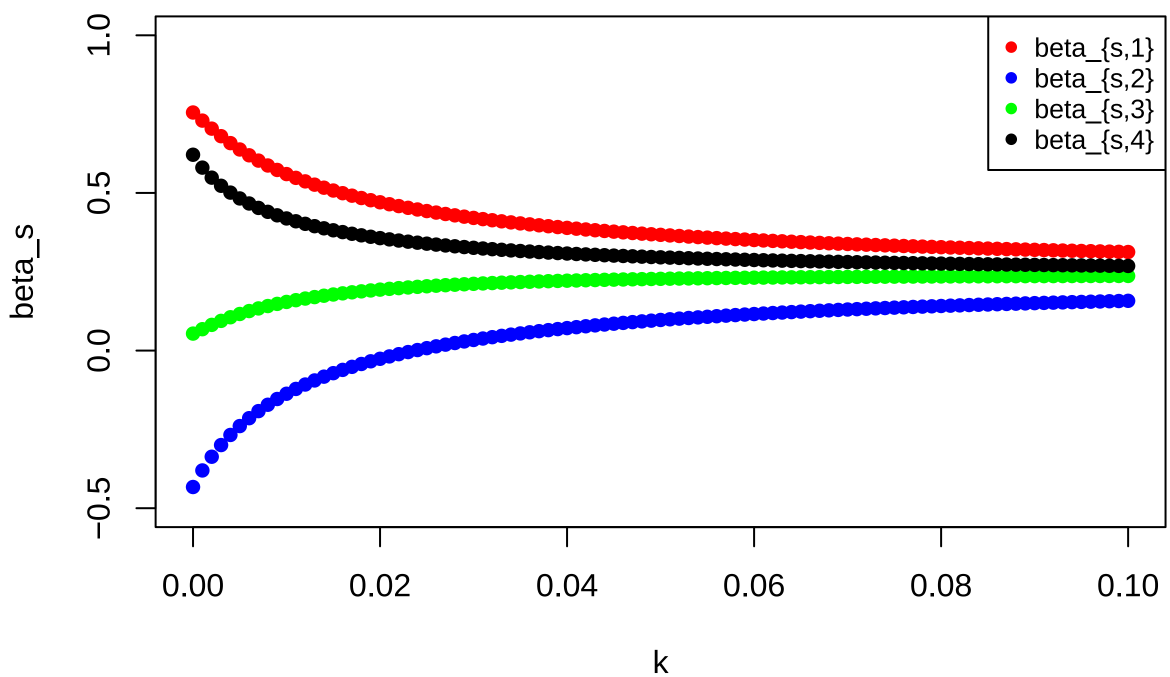
(a) 相关系数矩阵的特征值 $\lambda_4 = 0.0036$，且条件数 $k > 1000$，故存在严重的复共线性关系。

(b) 部分同学反应老师课件的第一种方法不能画出岭迹图，这里我们用课件的第二种方法。

```r
# 计算不同岭参数下的岭估计值，存储于 tab 矩阵中
tab <- matrix(0,nrow=101,ncol=5)
colnames(tab) <- c("k","beta_{s,1}","beta_{s,2}","beta_{s,3}","beta_{s,4}")
for (k in 1:101) {
```

```
  beta <- solve(R+(k-1)*diag(4)/1000)%*%t(Xs)%*%Ys
  tab[k,] <- c((k-1)/1000,beta)
}
```

```
# 绘制岭迹图
plot(tab[,1],tab[,2],type="n",xlab="k",ylab="beta_s",ylim=c(-0.5,1))
points(tab[,1],tab[,2],col="red",pch=16)
points(tab[,1],tab[,3],col="blue",pch=16)
points(tab[,1],tab[,4],col="green",pch=16)
points(tab[,1],tab[,5],col="black",pch=16)
legend("topright", legend = c("beta_{s,1}", "beta_{s,2}", "beta_{s,3}", "beta_{s,4}"),
       col = c( "red", "blue", "green","black"), pch = 16, cex = 0.8)
```



(c) 从岭迹图可以看出来，岭参数的取值范围应该是 0.06-0.08，这里我们取 0.06 作为岭参数，从 tab 矩阵中即可取出相应的岭估计值，

```
beta_sr <- tab[61,2:5]
beta_sr
```

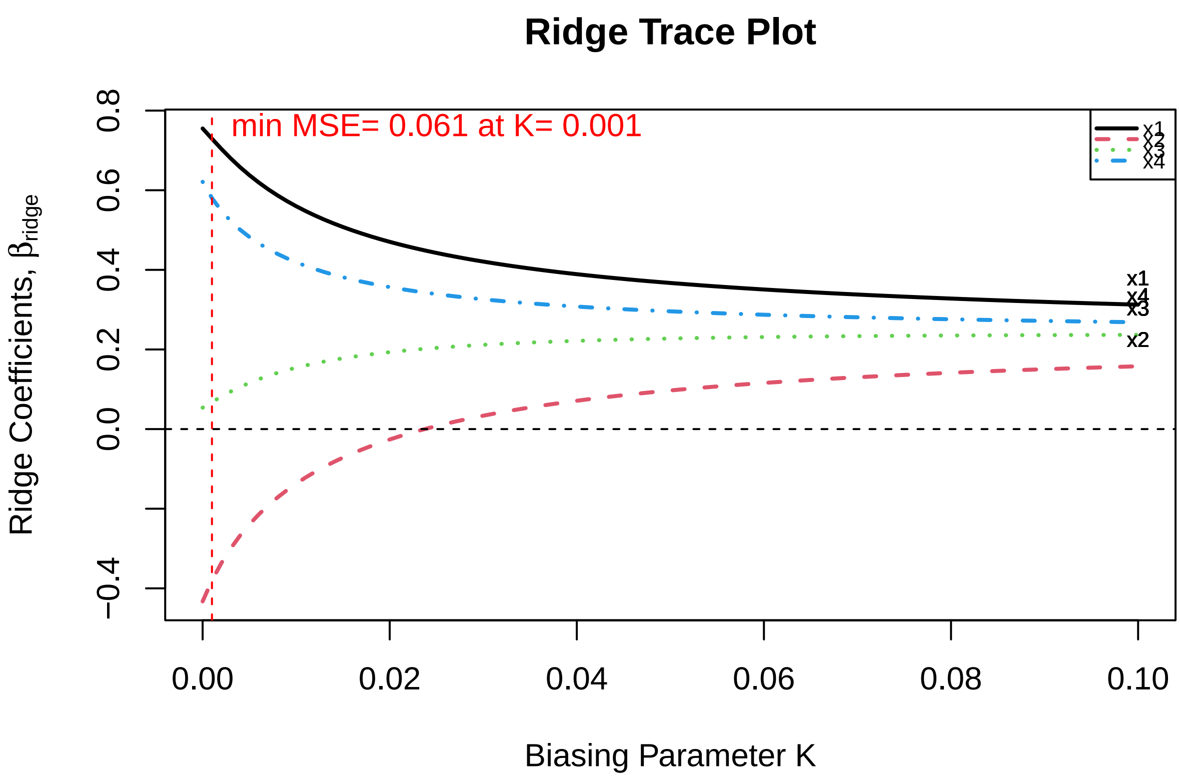
```
## beta_{s,1} beta_{s,2} beta_{s,3} beta_{s,4}
##  0.3506836  0.1158875  0.2311981  0.2872635
```

可得岭回归方程

$$\frac{Y-14.03}{13.03}=0.35\frac{X_1-129.37}{109.25}+0.12\frac{X_2-101.70}{24.37}+0.23\frac{X_3-36.35}{38.82}+0.29\frac{X_4-102.00}{19.24}$$

lmridge 包的函数也能求岭估计值，并且能直接画出岭迹图，用法如下。

```
d <- data.frame(Xs,Ys)
plot(lmridge(Ys~x1+x2+x3+x4,data=d,K=seq(0,0.1,0.001)))
```




```
lmridge(Ys~x1+x2+x3+x4,data=d,K=0.06)
```

```
## Call:
## lmridge.default(formula = Ys ~ x1 + x2 + x3 + x4, data = d, K = 0.06)
##
## Intercept         x1         x2         x3         x4
##    0.00000    0.35068    0.11589    0.23120    0.28726
```

(d) 利用公式 $\beta_{s,p}=\Phi_1\Phi_1^T\beta_s$ 求主成分估计值，

```
# 保留前三个主成分
phi <- eigen(R)$vec
beta_sp <- phi[,1:3]%*%t(phi[,1:3])%*%tab[1,2:5]
beta_sp
```

```
##            [,1]
## [1,]  0.8235953
## [2,] -0.3906552
## [3,]  0.0146640
## [4,]  0.5494720
```

可得主成分回归方程

$$\frac{Y-14.03}{13.03}=0.82\frac{X_1-129.37}{109.25}-0.39\frac{X_2-101.70}{24.37}+0.01\frac{X_3-36.35}{38.82}+0.55\frac{X_4-102.00}{19.24}$$

(e) 这是一个开放性问题，言之有理即可。下面提供一个例子，仅作参考。

首先，这组数据存在严重的共线性关系，因此不考虑 OLS 估计。关于岭估计与主成分估计，计算模型的残差平方和，选择残差平方和较小的模型。

```
# 岭估计的 RSS
sy^2*sum((Ys-Xs%*%beta_sr)^2)
```

```
## [1] 1.563737
```

```
# 主成分估计的 RSS
sy^2*sum((Ys-Xs%*%beta_sp)^2)
```

```
## [1] 0.215609
```

主成分估计的残差平方和更小，则测定系数更大，因此选择主成分估计。