HW9 STAT512 Fall2014

Yet Nguyen

December 4, 2014

1 & 2

```
# model matrix#####
library(plyr)
set.seed(1)
D <- matrix(c(-1, 1, 1, -1, 1, 1, 1, -1, -1, -1,
              -1, -1, -1, 1, -1, 1, 1, -1, 1, 1,
              1, -1, -1, -1, 1, -1, 1, 1, -1, 1,
              1, 1, -1, -1, -1, 1, -1, 1, 1, -1,
              -1, 1, 1, 1, -1, -1, -1, 1, -1, 1,
              1, -1, 1, 1, 1, -1, -1, -1, 1, -1),
            ncol = 10, byrow = T)
X <- cbind(rep(1, 6), D); intercept <- 2
m <- c(1, 2, 5, 10); s.b <- c(1, 2, 5, 10)
sim.func <- function(m, s.b){
  main.effect <- rep(0, 10)
  nonzero.effect<- sample(1:10, m)
  main.effect[nonzero.effect] <- rnorm(m, 0, s.b)
  noise <- rnorm(6, 0, 1)
  mean.vector <- c(intercept, main.effect)
  y <- X%*%(mean.vector) + noise
  Rsquare <- laply(2:11, function(i)
    summary(lm(y ~X[,i]))$r.squared)
  ans1 <- as.numeric(which.max(Rsquare) %in%
                       nonzero.effect)
  ans2 <- as.numeric(which.max(Rsquare) %in%
                       which.max(abs(nonzero.effect)))
  return(c(ans1=ans1, ans2 = ans2))
}

result.func <- function(m,s.b){
  out <- laply(1:1000, function(i)sim.func(m,s.b))
  res <- apply(out, 2, mean)
  return(res)
```

```
}

par <- matrix(c(1,1,
                1, 2,
                1,3,
                1,4,
                2,1,
                2,2,
                2,3,
                2,4,
                3,1,
                3,2,
                3,3,
                3,4,
                4,1,
                4,2,
                4,3,
                4,4), ncol = 2, byrow = T)
final.result <- laply(1:16, function(i)
  result.func(m[par[i,1]],s.b[par[i,2]]))
final.result2 <- cbind(m = as.character(rep(m, each = 4)),
                       sigma.beta = as.character(rep(s.b, 4)),
                       final.result)
```

	m	sigma.beta	ans1	ans2
1	1	1	0.461	0.097
2	1	2	0.656	0.086
3	1	5	0.875	0.104
4	1	10	0.918	0.092
5	2	1	0.628	0.108
6	2	2	0.823	0.098
7	2	5	0.936	0.096
8	2	10	0.975	0.078
9	5	1	0.778	0.098
10	5	2	0.853	0.099
11	5	5	0.852	0.101
12	5	10	0.869	0.084
13	10	1	1	0.093
14	10	2	1	0.108
15	10	5	1	0.101
16	10	10	1	0.098

Table 1: Simulation result based on 1000 replication for each scenario of m and sigma.beta. The column ans1 is the probability that the apparently best-fitting main effect is actually one of the active factor. The column ans2 is the probability that the apparently best-fitting main effect is actually the active factor with the largest absolute value main effect.