Simulation

```
# example 2 of the thesis
generate = function(n = 200, # number of observations
                    p = 200, # number of predictors
                    ps = 3, # number of strong signals
                    pwbc = 30, # number of WBC signals
                    pwai = 30, # number of WAI signals
                    corr = 0.7, # correlation coefficient (see example 2)
                    c = 20 # the c in the definition of strong and weak
                    ) {
  if (ps + pwbc + pwai >= n) {
    stop("number of true predictors should be less than number of observations")
  }
  if (abs(corr) > 1) {
    stop("correlation coefficient must be between -1 and 1")
  }
  # we hope that beta_strong = 20 and beta_weak = 0.5
  if ((c*sqrt(log(p)/n)) <= 0.5 & (c*sqrt(log(p)/n)) >= 20) {
    stop("please select other n and p")
  }
  corr_matrix = matrix(rep(0, len = p^2), nrow = p)
  corr_num = pwbc %/% ps
  # correlations between strong & wbc
  for (i in 1:(ps - 1)) {
    for (j in (ps + 1 + (i - 1)*corr_num):(ps + i*corr_num)) {
      corr_matrix[i, j] = corr
      corr_matrix[j, i] = corr
      for (k in j:(ps + i*corr_num)) {
        corr_matrix[j, k] = corr
        corr_matrix[k, j] = corr
      }
    }
  }
  for (j in (ps + 1 + (ps - 1)*corr_num):(ps + pwbc)) {
    corr_matrix[i, j] = corr
    corr_matrix[j, i] = corr
    for (k in j:(ps + pwbc)) {
      corr_matrix[j, k] = corr
      corr_matrix[k, j] = corr
    }
  }
  # correlations within wai
  for (j in (ps + pwbc + 1):(ps + pwbc + pwai)) {
    for (k in j:(ps + pwbc + pwai)) {
      corr_matrix[j, k] = corr
      corr_matrix[k, j] = corr
    }
```

```
  }
  diag(corr_matrix) = 1
  X = mvrnorm(n, mu = rep(0, p), Sigma = corr_matrix, tol = 1)
  beta = c(rep(20, ps), rep(0.5, pwbc + pwai), rep(0, p - ps - pwbc - pwai))
  Y = X %*% beta + rnorm(n)
  list_names = c("y",
                 paste("s", 1:ps, sep = "_"),
                 paste("wbc", 1:pwbc, sep = "_"),
                 paste("wai", 1:pwai, sep = "_"),
                 paste("null", 1:(p - ps - pwbc - pwai), sep = "_"))
  data = as.data.frame(cbind(Y, X))
  colnames(data)[1] <- "y"
  for (i in 2:(1 + ps)) {
    colnames(data)[i] <- paste("s", i - 1, sep = "_")
  }
  for (i in (2 + ps):(1 + ps + pwbc)) {
    colnames(data)[i] <- paste("wbc", i - 1 - ps, sep = "_")
  }
  for (i in (2 + ps + pwbc):(1 + ps + pwbc + pwai)) {
    colnames(data)[i] <- paste("wai", i - 1 - ps - pwbc, sep = "_")
  }
  for (i in (2 + ps + pwbc + pwai):p) {
    colnames(data)[i] <- paste("null", i - 1 - ps - pwbc - pwai, sep = "_")
  }
  return(data)
}
```

```
# retrieve data
data = generate()
df = data.frame(data)
n = 200

# Forward Selection
fit_forward = step(object = lm(y ~ 1, data = df),
                   scope = formula(lm(y ~ ., data = df)),
                   direction = "forward",
                   k = 2*n,
                   trace = 0)
summary(fit_forward)
```

```
##
## Call:
## lm(formula = y ~ 1, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -122.010  -31.912    2.924   30.477  110.671
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.406      2.972  -0.473    0.637
##
## Residual standard error: 42.03 on 199 degrees of freedom
```

```r
# LASSO
X = as.matrix(data[2:n + 1])
Y = as.matrix(data[1])
fit_lasso = cv.glmnet(X, Y, nfolds = 10, type.measure = "mse")
param.best = fit_lasso$glmnet.fit$beta[, fit_lasso$lambda == fit_lasso$lambda.1se]
param.best
```

```
##         s_2         s_3       wbc_1       wbc_2       wbc_3       wbc_4
## 22.27692015 18.16185664  4.17326982  0.00000000  1.04021605  1.67275454
##       wbc_5       wbc_6       wbc_7       wbc_8       wbc_9      wbc_10
##  0.64560307  5.40336974  2.04787239  1.38428875  1.57131818  2.00732512
##      wbc_11      wbc_12      wbc_13      wbc_14      wbc_15      wbc_16
##  0.00000000  0.00000000  0.00000000  0.04294236  0.00000000  0.00000000
##      wbc_17      wbc_18      wbc_19      wbc_20      wbc_21      wbc_22
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##      wbc_23      wbc_24      wbc_25      wbc_26      wbc_27      wbc_28
##  0.62002143  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##      wbc_29      wbc_30       wai_1       wai_2       wai_3       wai_4
##  0.00000000  0.00000000  0.13116469  0.00000000  0.00000000  2.47426048
##       wai_5       wai_6       wai_7       wai_8       wai_9      wai_10
##  1.22959767  0.00000000  0.11098182  0.00000000  0.00000000  0.00000000
##      wai_11      wai_12      wai_13      wai_14      wai_15      wai_16
##  0.52406902  0.00000000  0.00000000  0.00000000  0.95305119  0.00000000
##      wai_17      wai_18      wai_19      wai_20      wai_21      wai_22
##  0.00000000  0.00000000  0.00000000  0.56816883  1.05370721  0.00000000
##      wai_23      wai_24      wai_25      wai_26      wai_27      wai_28
##  0.00000000  0.00000000  0.00000000  3.01450452  0.00000000  0.00000000
##      wai_29      wai_30      null_1      null_2      null_3      null_4
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##      null_5      null_6      null_7      null_8      null_9     null_10
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_11     null_12     null_13     null_14     null_15     null_16
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_17     null_18     null_19     null_20     null_21     null_22
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_23     null_24     null_25     null_26     null_27     null_28
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_29     null_30     null_31     null_32     null_33     null_34
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_35     null_36     null_37     null_38     null_39     null_40
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_41     null_42     null_43     null_44     null_45     null_46
##  0.00000000  0.00000000  0.00000000 -0.40449752  0.00000000  0.00000000
##     null_47     null_48     null_49     null_50     null_51     null_52
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_53     null_54     null_55     null_56     null_57     null_58
##  0.00000000  0.03876890  0.00000000  0.00000000  0.00000000  0.00000000
##     null_59     null_60     null_61     null_62     null_63     null_64
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_65     null_66     null_67     null_68     null_69     null_70
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_71     null_72     null_73     null_74     null_75     null_76
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_77     null_78     null_79     null_80     null_81     null_82
```

```
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_83     null_84     null_85     null_86     null_87     null_88
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_89     null_90     null_91     null_92     null_93     null_94
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##     null_95     null_96     null_97     null_98     null_99    null_100
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_101    null_102    null_103    null_104    null_105    null_106
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_107    null_108    null_109    null_110    null_111    null_112
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_113    null_114    null_115    null_116    null_117    null_118
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_119    null_120    null_121    null_122    null_123    null_124
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_125    null_126    null_127    null_128    null_129    null_130
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##    null_131    null_132    null_133    null_134    null_135    null_136
##  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
##        V201
##  0.00000000
```