Simulation

```r
# Example 2 of the thesis
generate = function(n, # number of observations
                    p, # 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)
  df = as.data.frame(cbind(Y, X))
  colnames(df)[1] <- "y"
  for (i in 2:(1 + ps)) {
    colnames(df)[i] <- paste("strong", i - 1, sep = "_")
  }
  for (i in (2 + ps):(1 + ps + pwbc)) {
    colnames(df)[i] <- paste("wbc", i - 1 - ps, sep = "_")
  }
  for (i in (2 + ps + pwbc):(1 + ps + pwbc + pwai)) {
    colnames(df)[i] <- paste("wai", i - 1 - ps - pwbc, sep = "_")
  }
  for (i in (2 + ps + pwbc + pwai):(1 + p)) {
    colnames(df)[i] <- paste("null", i - 1 - ps - pwbc - pwai, sep = "_")
  }
  return(df)
}
```

```
# Number of observations
n = 1000
# Number of predictors
p = 200

# Data manipulation
df = generate(n, p)
X = as.matrix(df[1:p + 1])
Y = as.matrix(df[1])

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

```
##
## Call:
## lm(formula = y ~ strong_2 + strong_1 + strong_3 + wai_1 + wai_12 +
##     wai_20 + wai_17 + wai_18 + wbc_8 + wai_27 + wai_22 + wai_23 +
##     wbc_5 + wai_9 + wai_26 + wai_13 + wbc_3 + wai_4 + wai_25 +
##     wbc_2 + wai_11 + wai_24 + wai_10 + wbc_9 + wai_15 + wbc_1 +
##     wai_3 + wai_30 + wbc_7 + wai_2 + wai_8 + wai_29 + wbc_10 +
##     wai_19 + wai_21 + wai_6 + wai_7 + wbc_6 + wai_14 + wai_5 +
##     wai_28 + wai_16 + wbc_4 + null_4 + null_133 + null_32 + null_19 +
##     null_20 + null_17 + null_41 + null_78 + null_77 + wbc_13 +
##     null_87 + null_115 + null_68 + null_66 + null_111 + null_29 +
##     null_30 + null_99 + null_120 + null_91 + null_52 + null_98 +
##     null_90 + null_12 + wbc_22 + null_46 + null_9, data = df)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.07408 -0.64076  0.01612  0.66739  3.10227
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.007722   0.032167  -0.240 0.810342
## strong_2    25.295489   0.053955 468.829  < 2e-16 ***
## strong_1    20.009386   0.057698 346.794  < 2e-16 ***
## strong_3    20.001008   0.032731 611.077  < 2e-16 ***
## wai_1        0.401739   0.059137   6.793 1.95e-11 ***
## wai_12       0.635099   0.057610  11.024  < 2e-16 ***
## wai_20       0.542396   0.054782   9.901  < 2e-16 ***
## wai_17       0.634601   0.056204  11.291  < 2e-16 ***
## wai_18       0.479990   0.057422   8.359 2.29e-16 ***
## wbc_8        0.676563   0.054999  12.301  < 2e-16 ***
## wai_27       0.613699   0.059088  10.386  < 2e-16 ***
## wai_22       0.492599   0.056626   8.699  < 2e-16 ***
## wai_23       0.480641   0.056631   8.487  < 2e-16 ***
## wbc_5        0.411673   0.061168   6.730 2.96e-11 ***
## wai_9        0.490152   0.056567   8.665  < 2e-16 ***
## wai_26       0.629849   0.055654  11.317  < 2e-16 ***
## wai_13       0.516524   0.057551   8.975  < 2e-16 ***
## wbc_3        0.525768   0.056827   9.252  < 2e-16 ***
## wai_4        0.437596   0.057018   7.675 4.19e-14 ***
## wai_25       0.546339   0.055745   9.801  < 2e-16 ***
## wbc_2        0.590524   0.057704  10.234  < 2e-16 ***
## wai_11       0.577421   0.055997  10.312  < 2e-16 ***
## wai_24       0.465604   0.057651   8.076 2.06e-15 ***
## wai_10       0.521127   0.057877   9.004  < 2e-16 ***
## wbc_9        0.528853   0.056047   9.436  < 2e-16 ***
## wai_15       0.523171   0.056172   9.314  < 2e-16 ***
## wbc_1        0.431697   0.054348   7.943 5.67e-15 ***
## wai_3        0.541725   0.056938   9.514  < 2e-16 ***
## wai_30       0.517091   0.057789   8.948  < 2e-16 ***
## wbc_7        0.475720   0.055825   8.522  < 2e-16 ***
## wai_2        0.506553   0.058338   8.683  < 2e-16 ***
## wai_8        0.407771   0.058822   6.932 7.74e-12 ***
## wai_29       0.470951   0.057996   8.120 1.47e-15 ***
## wbc_10       0.474960   0.056241   8.445  < 2e-16 ***
## wai_19       0.475889   0.055307   8.604  < 2e-16 ***
## wai_21       0.463898   0.058200   7.971 4.61e-15 ***
## wai_6        0.466351   0.055845   8.351 2.45e-16 ***
## wai_7        0.470673   0.056429   8.341 2.64e-16 ***
## wbc_6        0.508014   0.056659   8.966  < 2e-16 ***
## wai_14       0.496141   0.059367   8.357 2.33e-16 ***
## wai_5        0.410162   0.057236   7.166 1.57e-12 ***
## wai_28       0.433474   0.059279   7.312 5.66e-13 ***
## wai_16       0.396260   0.058253   6.802 1.84e-11 ***
## wbc_4        0.361506   0.055073   6.564 8.69e-11 ***
## null_4       0.113890   0.031928   3.567 0.000379 ***
## null_133     0.106999   0.032228   3.320 0.000935 ***
## null_32      0.108137   0.032420   3.335 0.000885 ***
```

```
## null_19     -0.104419   0.032099  -3.253 0.001183 **
## null_20      0.081789   0.031861   2.567 0.010411 *
## null_17     -0.088082   0.031955  -2.756 0.005958 **
## null_41      0.079033   0.032349   2.443 0.014746 *
## null_78     -0.086122   0.031661  -2.720 0.006648 **
## null_77      0.076606   0.031872   2.404 0.016434 *
## wbc_13       0.138575   0.047775   2.901 0.003813 **
## null_87      0.067510   0.032833   2.056 0.040045 *
## null_115     0.073677   0.031980   2.304 0.021452 *
## null_68     -0.064951   0.032653  -1.989 0.046982 *
## null_66      0.062138   0.029636   2.097 0.036292 *
## null_111    -0.050826   0.031548  -1.611 0.107509
## null_29     -0.058981   0.031320  -1.883 0.059990 .
## null_30     -0.052218   0.031538  -1.656 0.098117 .
## null_99      0.059342   0.031502   1.884 0.059909 .
## null_120    -0.057090   0.031631  -1.805 0.071421 .
## null_91     -0.055598   0.033354  -1.667 0.095865 .
## null_52     -0.050285   0.033033  -1.522 0.128281
## null_98     -0.048739   0.031200  -1.562 0.118594
## null_90      0.047010   0.030621   1.535 0.125070
## null_12     -0.048739   0.031757  -1.535 0.125186
## wbc_22       0.078191   0.048296   1.619 0.105788
## null_46     -0.048111   0.031580  -1.523 0.127982
## null_9      -0.047278   0.031808  -1.486 0.137515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9758 on 929 degrees of freedom
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9995
## F-statistic: 3.145e+04 on 70 and 929 DF,  p-value: < 2.2e-16
```

```
params_forward = fit_forward$coefficients[-1]
params_forward
```

```
##    strong_2    strong_1    strong_3       wai_1      wai_12      wai_20
## 25.29548861 20.00938587 20.00100828  0.40173914  0.63509941  0.54239595
##      wai_17      wai_18       wbc_8      wai_27      wai_22      wai_23
##  0.63460104  0.47999028  0.67656317  0.61369859  0.49259906  0.48064118
##       wbc_5       wai_9      wai_26      wai_13       wbc_3       wai_4
##  0.41167291  0.49015160  0.62984926  0.51652406  0.52576780  0.43759612
##      wai_25       wbc_2      wai_11      wai_24      wai_10       wbc_9
##  0.54633943  0.59052412  0.57742149  0.46560369  0.52112718  0.52885333
##      wai_15       wbc_1       wai_3      wai_30       wbc_7       wai_2
##  0.52317130  0.43169706  0.54172468  0.51709146  0.47572033  0.50655298
##       wai_8      wai_29      wbc_10      wai_19      wai_21       wai_6
##  0.40777084  0.47095149  0.47495990  0.47588936  0.46389793  0.46635070
##       wai_7       wbc_6      wai_14       wai_5      wai_28      wai_16
##  0.47067295  0.50801433  0.49614075  0.41016158  0.43347407  0.39626006
##       wbc_4      null_4    null_133     null_32     null_19     null_20
##  0.36150568  0.11389003  0.10699859  0.10813701 -0.10441898  0.08178900
##     null_17     null_41     null_78     null_77      wbc_13     null_87
## -0.08808231  0.07903326 -0.08612236  0.07660582  0.13857490  0.06751004
##    null_115     null_68     null_66    null_111     null_29     null_30
##  0.07367699 -0.06495149  0.06213767 -0.05082599 -0.05898074 -0.05221774
```

```
##     null_99    null_120     null_91     null_52     null_98     null_90
##  0.05934190 -0.05709011 -0.05559848 -0.05028548 -0.04873925  0.04701043
##     null_12      wbc_22     null_46      null_9
## -0.04873856  0.07819059 -0.04811093 -0.04727846
```

```
# LASSO
fit_lasso = cv.glmnet(X, Y,
                      nfolds = 10,
                      type.measure = "mse")
param_lasso = fit_lasso$glmnet.fit$beta[, fit_lasso$lambda == fit_lasso$lambda.1se]
param_lasso
```

```
##   strong_1   strong_2   strong_3      wbc_1      wbc_2      wbc_3      wbc_4
## 20.0052585 25.0354613 19.5689981  0.3141667  0.5464197  0.4739149  0.2594344
##      wbc_5      wbc_6      wbc_7      wbc_8      wbc_9     wbc_10     wbc_11
##  0.4793141  0.3518509  0.3719128  0.6216419  0.4699600  0.4502717  0.0000000
##     wbc_12     wbc_13     wbc_14     wbc_15     wbc_16     wbc_17     wbc_18
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##     wbc_19     wbc_20     wbc_21     wbc_22     wbc_23     wbc_24     wbc_25
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##     wbc_26     wbc_27     wbc_28     wbc_29     wbc_30      wai_1      wai_2
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.2884951  0.4941180
##      wai_3      wai_4      wai_5      wai_6      wai_7      wai_8      wai_9
##  0.5547449  0.5403319  0.4495702  0.4451067  0.4257245  0.4356972  0.3656373
##     wai_10     wai_11     wai_12     wai_13     wai_14     wai_15     wai_16
##  0.4888274  0.4913452  0.6501002  0.5062421  0.4901149  0.4835426  0.3058103
##     wai_17     wai_18     wai_19     wai_20     wai_21     wai_22     wai_23
##  0.6136230  0.4719756  0.5104335  0.5152844  0.5240354  0.5285334  0.4928308
##     wai_24     wai_25     wai_26     wai_27     wai_28     wai_29     wai_30
##  0.4182108  0.4649379  0.6231632  0.5401755  0.3842580  0.4920385  0.4629532
##     null_1     null_2     null_3     null_4     null_5     null_6     null_7
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##     null_8     null_9    null_10    null_11    null_12    null_13    null_14
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_15    null_16    null_17    null_18    null_19    null_20    null_21
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_22    null_23    null_24    null_25    null_26    null_27    null_28
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_29    null_30    null_31    null_32    null_33    null_34    null_35
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_36    null_37    null_38    null_39    null_40    null_41    null_42
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_43    null_44    null_45    null_46    null_47    null_48    null_49
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_50    null_51    null_52    null_53    null_54    null_55    null_56
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_57    null_58    null_59    null_60    null_61    null_62    null_63
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_64    null_65    null_66    null_67    null_68    null_69    null_70
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_71    null_72    null_73    null_74    null_75    null_76    null_77
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_78    null_79    null_80    null_81    null_82    null_83    null_84
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
```

```
##    null_85    null_86    null_87    null_88    null_89    null_90    null_91
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_92    null_93    null_94    null_95    null_96    null_97    null_98
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##    null_99   null_100   null_101   null_102   null_103   null_104   null_105
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##   null_106   null_107   null_108   null_109   null_110   null_111   null_112
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##   null_113   null_114   null_115   null_116   null_117   null_118   null_119
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##   null_120   null_121   null_122   null_123   null_124   null_125   null_126
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##   null_127   null_128   null_129   null_130   null_131   null_132   null_133
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##   null_134   null_135   null_136   null_137
##  0.0000000  0.0000000  0.0000000  0.0000000
```