Statistical Learning - Report 1

Marta Kałużna

4/3/2020

Exercise 1

For a given matrix A and vector b we're going to show that $x^* = (1, 1, -1, 0)$ satisfies the irrepresantable condition but is not one of the sparsest solutions of $Ax = b$.

When $x_1 = 0$, the solution for given equation is $\widetilde{x} = (0, 0, -2, \sqrt{3})$, which is sparsest than x^* ($||x^*||_0 > ||\widetilde{x}||_0$). In the same time, x^* satisfies the irrepresentable condition:

```
x_star <- c(1, 1, -1, 0)
not_supp_x <- which(x_star == 0)
supp_x <- which(x_star != 0)

A <- matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 1, 1/sqrt(3), 1/sqrt(3), 1/sqrt(3)), nrow = 3, ncol = 4)
A_supp <- A[,supp_x]
A_prim <- A[,not_supp_x]

max(abs(t(A_prim) %*% A_supp %*% solve(t(A_supp) %*% A_supp) %*% sign(x_star[supp_x]))) < 1
```

```
## [1] TRUE
```

Exercise 2

We're going to work on a particular observation B of a 200x500 Gaussian random matrix with i.i.d.$N(0, 1)$ entries. A is a normalization of B.

1) Firstly, we will compute mutual coherence.

```
set.seed(2020)
n <- 200
p <- 500
B <- matrix(rnorm(n*p), nrow = n, ncol = p)
A <- apply(B, 2, function(x) x/sqrt(sum(x^2)))

mutual_coherence <- function(norm_matrix, n, p){
  dot_product_vector <- c()
  k <- 1
  for(i in 1:p){
    dot_vec <- as.vector(norm_matrix[,i])
    for(j in i:p){
      if(i != j){
      dot_product_vector[k] <- abs(as.vector(norm_matrix[,j]) %*% dot_vec)
      k <- k + 1
      }
    }
  }
  return(max(dot_product_vector))
}
```

```
MA <- mutual_coherence(A, n, p)
```

$M(A) =$ 0.3021009

Then, we will provide the set K_0 of sparsity $k \in \{1, ..., 200\}$ for which, according to *mutual coherence condition*, x^* is both the unique basis pursuit minimizer and the unique sparsest solution of Ax = b.

```
mutual_coherence_cond <- function(){
  cond_true <- c()
  for(k in 1:200){
    cond_true[k] <- (k < (1 + (1/MA)) / 2)
  }
 return (cond_true)
}

K_0 <- which(mutual_coherence_cond() == 1)
```

$K_0 = \{$1, 2$\}$
Because the condition is very strict and $M(A)$ in our case is really small, it's hard to satisfy this condition. That's why the set K_0 is small.

2) In the second part, we'll provide the set K_1 of sparsity $k \in \{1, ..., 200\}$ for which x^* satisfies *the irrepresentable condition.*

```
irrepresentable_condition <- function(p){
  cond_true <- c()
  for(k in 1:200){
    x_star <- c(abs(rnorm(k)), rep(0, times = p - k))
    not_supp_x <- which(x_star == 0)
    supp_x <- which(x_star != 0)
    A_supp <- A[,supp_x]
    A_prim <- A[,not_supp_x]
    cond_true[k] <- max(abs(t(A_prim) %*% A_supp %*% solve(t(A_supp) %*% A_supp)
                        %*% sign(x_star[supp_x]))) < 1
  }
  return (cond_true)
}

K_1 <- which(irrepresentable_condition(500) == 1)
```

$K_1 = \{$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23$\}$

This time, we have much more elements in the set, because the condition is less strict.

3)

```
irr_cond <- function(x, matrix){
  not_supp_x <- which(x == 0)
  supp_x <- which(x != 0)
  mat_supp <- matrix[,supp_x]
  mat_prim <- matrix[,not_supp_x]
  held_or_not <- max(abs(t(mat_prim) %*% mat_supp %*% solve(t(mat_supp) %*% mat_supp)
                     %*% sign(x[supp_x]))) < 1
  return(held_or_not)
}

zad3 <- function(matrix) {
```

```r
  x_irr_vec <- c()
  x_wave_irr_vec <- c()
  for(k in 1:200){
    x <- c(abs(rnorm(k)), rep(0, 500 - k))
    x_wave <- c(rep(1, k), rep(0, 500 - k))
    x_irr_vec[k] <- irr_cond(x, A)
    x_wave_irr_vec[k] <- irr_cond(x_wave, A)
  }
  compare_matrix <- matrix(0,ncol = 2,nrow = length( x_irr_vec) )
  compare_matrix[,1] <- x_irr_vec
  compare_matrix[,2] <- x_wave_irr_vec
  return(compare_matrix)
}

compare <- zad3(A)
compare[1:30,]
```

```
##       [,1] [,2]
##  [1,]    1    1
##  [2,]    1    1
##  [3,]    1    1
##  [4,]    1    1
##  [5,]    1    1
##  [6,]    1    1
##  [7,]    1    1
##  [8,]    1    1
##  [9,]    1    1
## [10,]    1    1
## [11,]    1    1
## [12,]    1    1
## [13,]    1    1
## [14,]    1    1
## [15,]    1    1
## [16,]    1    1
## [17,]    1    1
## [18,]    1    1
## [19,]    1    1
## [20,]    1    1
## [21,]    1    1
## [22,]    1    1
## [23,]    1    1
## [24,]    0    0
## [25,]    0    0
## [26,]    0    0
## [27,]    0    0
## [28,]    0    0
## [29,]    0    0
## [30,]    0    0
```

First column says if x^* is the only solution of the BP minimizer. The second one is the same but for $\widetilde{x}$. We can easily see that x^* satisifies irrepresantable condition if and only if $\widetilde{x}$ does.

4) Provide the set K2 of sparsity $k \in \{1, ..., 200\}$ for which x^* is *the BP minimizer* of Ax = b.

```r
library(ADMM)

improveBP = function(A,b,x){
  n <- nrow(A)
  p <- ncol(A)
  u <- order(abs(x))
  J <- u[1:(p-n)]
  x0 <- rep(0, p)
  A1 <- A[,-J]
  v <- solve(A1) %*% b
  x0[-J] <- v
  return(x0)
}

bp <- c()
for(k in (1:n)){
  x0 <- c(rep(1, k), rep(0, p - k))
  b <- A%*%x0
  BPminimizer <- admm.bp(A, b, abstol = 0.00001, reltol = 0.00001, maxiter = 10000)$x
  BPbetter <- improveBP(A, b, BPminimizer)
  bp[k] <- (max(abs(BPbetter - x0)) <= 0.000001)
}

K_2 <- which(bp == 1)
```

K_2 = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63}
K_2 is the biggest set. Checking for which sparsity x^* is a BP minimizer is the weakest condition through all previously checked. Because there is no assumption about the singleton we expect this set to be larger.

5) Provide the set K3 of sparsity $k \in \{1, ..., 200\}$ for which x^* is *both the unique BP minimizer and the unique sparsest solution* of Ax = b.

From mutual coherence condition, we know that x^* is both the unique BP minimizer and the unique sparsest solution of Ax = b when $||x^*||_0 < (1 + 1/M(A))/2$. As we could see, it is a very strict condition and we may lose some information.
There are two independent assumptions for BP and sparsest solution. We will use both of them for each x^*. Then we will compare K_3 with K_0.

To find sparsest solution we will calculate $spark(A)$. Then we'll find out if $||x||_0 < spark(A)/2$ which implicates x^* is the unique sparsest solution. For checking if x^* is the unique BP minimizer, we will use irrepresentable condition.

```r
library(pracma)

spark <- function(matrix) {
  kernel <- nullspace(A)
  kernel[abs(kernel) < 0.01] <- 0
  min_spark <- dim(kernel)[1] - max(colSums(kernel == 0))
  return(min_spark)
}

spark_number <- spark(A)

zad5 <- function(matrix){
```

```r
  is_bp_sp <- c()
  for(k in 1:200) {
    x <- c(rep(1, k), rep(0, 500 - k))
    is_bp <- irr_cond(x,matrix)
    is_sp <- (sum(x != 0) < spark_number/2)
    is_bp_sp[k] <- is_bp & is_sp
  }
  return(is_bp_sp)
}

K_3 <- which(zad5(A) == 1)
```

$K_3 = \{$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23$\}$
Set K_3 has much more elements than set K_0. As I said before, mutual coherence condition is very strict and can cause many false negatives.

Exercise 6

The last exercise shows that the mutual coherence condition implies the irrepresentable condition.

At the begining we need to create a diagonal dominant matrix Q.

```r
Q <- matrix(rnorm(500*500), ncol = 500, nrow  = 500)

for (i in 1:500) {
  Q[i,i] <- (sum(abs(Q[i, -i])) + abs(rnorm(1)))*(-1)^sample(1:100,1)
}

nullspace(Q)
```

```
## NULL
```

We just checked that $ker(Q) = 0$.

Next step is to check Gershgorin circle theorem. To do that we need to show that each eigenvalue of Q belongs to the intervals constructed by subtracting/adding to the values on diagonal all the values that are in the same row (wihout value on the diagonal).

```r
G_circle <- function(d){
  lambdas <- Re(eig(d))
  is_between <- rep(0, 500)
  for (i in 1:500){
    lambda <- lambdas[i]
    is_this_one <- c()
      for(j in 1:500) {
        left_side <- Q[j,j] - sum(abs(d[j,-j]))
        right_side <- Q[j,j] + sum(abs(d[j,-j]))
        is_this_one[j] <- (lambda >= left_side) & (lambda <= right_side)
        }
    if(length(is_this_one)>=1) is_between[i] <- TRUE
  }
  return(is_between)
}

interval_check <- G_circle(Q)
sum(interval_check) == 500 # if it's true, condition is satisfied
```

```
## [1] TRUE
```

We are assuming that $||x^*||_0 \leq (1 + 1/M(A))/2$, so in our case left side is equal to 2.155.

We should prove that the largest eigenvalue of $(A'A)^{-1}$, where A are the only columns that are in the $supp(x^*)$, is smaller than $2/(M(A)+1)$. We will check for which level of sparsity from 1:200 this condition is fulfilled:

```
zad6 <- function(MA, matrix){
  is_cond_held <- c()
  for(k in 1:200){
     max_eigen <- max(eig(solve(t(A[,1:k]) %*% A[,1:k])))
     is_cond_held[k] <- max_eigen < 2/(MA + 1)
  }
  return(is_cond_held)
}

smaller_or_not <- zad6(MA,A)
```

This set is: $\{$`1, 2, 3, 4, 5, 6, 7, 8`$\}$. Thus, knowing the output, we can go beyond our start assumption that $||x^*||_0 < 2$.
We should then conclude that mutual coherence condition implies the irrepresentable condition. We could think about it when we were comparing sizes of the "K-sets". K_0 was the smallest one and was included in the others.

Now I will use some explanation according to paper on the website: For the next part we assume that $k = ||x^*||$ and $Q = A_I^T A_I$ where I are indexes that are in $supp(x^*)$. We know that diagonal elements of Q are ones and non-diagonal elements are smaller than M(A). By assumption M(A) is the biggest scalar product between the pair of different columns. Proving the Gerhgorin theroem we know that each eigenvalue λ of Q is in such interval (we approximate it):

`[1-(k-1)M(A), 1+(k-1)M(A)]`

Eigenvalues for $(A_I^T A_I)^{-1}$ wil be $\frac{1}{\lambda}$. Using previous interval we know that $\frac{1}{\lambda} < \frac{1}{1-M(A)k+M(A)} < \frac{2}{M(A)+1}$. Thus, we did point ii).