Lab1 Group11

Sofie Jörgensen, Oriol Garrobé Guilera, David Hrabovszki

1/24/2020

Question 1

1.

```
## [1] "Subtraction is wrong"
```

```
## [1] "Subtraction is correct"
```

After running the two code snippets given in the lab pdf, we observe that the first snippet prints "Subtraction is wrong", while the second snippet prints "Subtraction is correct". In both cases they should be mathematically correct, but because floats are rounded in R, the values 1/3 and 1/12 are not precisely represented. Since the two sides are not precisely equal, this will return "Subtraction is wrong".

2.

The problems mentioned in the 1. can be solved by using the functions all.equal() and isTRUE(), where all.equal() tests if two objects are (nearly) equal, in contrast with the == sign which tests perfect equality, and isTRUE() is necessary to handle the FALSE result of all.equal().

The following code snippet prints "Subtraction is correct", which is the correct output.

```
#1.2

x1<-1/3
x2<-1/4
if (isTRUE(all.equal(x1-x2,1/12))) {
print ( "Subtraction is correct" )
} else {
print ( "Subtraction is wrong")
}
```

```
## [1] "Subtraction is correct"
```

Question 2

Using the definition of a derivative, we use the following formula at point x with small ϵ:

$f'(x) = \frac{f(x+\epsilon)-f(x)}{\epsilon}$

1.

We calculate the derivative of $f(x) = x$, with $\epsilon = 10^{-15}$ using our own function (see Appendix).

2. & 3.

Now we evaluate our derivative function at $x = 1$ and $x = 100000$ and obtain the following results:

```
## [1] 1.110223
```

```
## [1] 0
```

The true values of the derivatives are 1 for every x, and we expect our function to output a value close to 1.

The obtained derivative for $x = 1$ is close to the true value, but for large x we get the result 0.

If we consider x within the range [0,1[, we obtain approximately 1. The difference between large numbers, dominates the small epsilon, which gives us zero in the nominator and thus the derivative evaluated at large points is 0. In general, adding small numbers first will give a better accuracy, than adding large number first, which we are doing when using the definition of the derivative.

Question 3

A known formula for estimating the variance based on a vector of n observations is

$$Var(\vec{x}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i \right)^2 \right)$$

1.

Write your own R function, myvar, to estimate the variance in this way.

2.

Generate a vector $x = (x_1, ..., x_{10000})$ with 10000 random numbers with mean 10^8 and variance 1.

3.

For each subset $X_i = \{x_1, ..., x_i\}, i = 1, ..., 10000$ compute the difference $Y_i = myvar(X_i) - var(X_i)$, where $var(X_i)$ is the standard variance estimation function in R.

4.

Question 4

1.

Import the data set to R.

```
#4.1
```

2.

Optimal regression coefficients can be found by solving a system of the type $A\vec{\beta} = \vec{b}$ where $A = X^T X$ and $\vec{b} = X^T \vec{y}$. The matrix X are the observations of the absorbance records, levels of moisture and fat, while $\vec{y}$ are the protein levels.

```
#4.2
```

3.

Try to solve $A\vec{\beta} = \vec{b}$ with default solver solve().

```
#4.3
```

4.

Check the condition number of the matrix A.

The condition number κ of square matrix A is: $||A||\,||A^{-1}||$

```
#4.4
```

5.

Scale the data set and repeat steps 2-4.

```
#4.5
```

Appendix

```r
knitr::opts_chunk$set(echo = TRUE)
RNGversion('3.5.1')

# Seed
set.seed(1234567890)

#1.1

x1<-1/3
x2<-1/4
if (x1-x2==1/12) {
print ( "Subtraction is correct" )
} else {
print ( "Subtraction is wrong")
}
x1<-1
x2<-1/2
if (x1-x2 ==1/2) {
  print("Subtraction is correct")
} else {
  print("Subtractino is wrong")
}
#1.2

x1<-1/3
x2<-1/4
if (isTRUE(all.equal(x1-x2,1/12))) {
print ( "Subtraction is correct" )
} else {
print ( "Subtraction is wrong")
}
#2.1

f <- function(x) x

epsilon <- 1e-15

derivative <- function(f, x, e){
  (f(x+epsilon) - f(x))/epsilon
```

```
}


#2.2 & 2.3

derivative(f, 1, epsilon)
derivative(f,100000, epsilon)
#4.1


#4.2


#4.3

#4.4

#4.5
```