LINKÖPINGS UNIVERSITET
Department of information and computer science
Statistics and Machine learning
Måns Magnusson

Exam

R programming (732G36)

Time:	8-12, 2015-10-14
Material:	The extra material is included in the zip-file **exam_material.zip**.
Grades:	A = 19-20 points.
	B = 17-18 points.
	C = 12-16 points.
	D = 10-11 points.
	E = 8-9 points.
	F = 0-7 points.

Instructions

Write your code in an R script file named **Main.R**. The R code should be complete and readable code, possible to run by copying directly into a script. Comment directly in the code whenever something needs to be explained or discussed. Follow the instructions carefully.

Problem 1 (6 p)

a) Create the following mathematical function as a function in R (called `f`)

$$f(\mathbf{x}) = \frac{\sum_i^n |x_i - \bar{x}|}{n}$$

where $\bar{x}$ is $\frac{1}{n}\sum_i x_i$ an n is the length of the vector $\mathbf{x}$. You are not allowed to use any vectorized functions such as `sum()` or `mean()`.

```
f(1:5)
```

```
[1] 1.2
```

```
f(c(7,2,2,1,-4))
```

```
[1] 2.48
```

b) What is the computational complexity of this algorithm based on the input length?

c) Visualize the value of $f(\mathbf{x})$ as $\mathbf{x}$ is growing by drawing 1,2,3,...,200 draws from a $\mathcal{N}(0,1)$ distribution and visualize the value of $f(\mathbf{x})$ using a linegraph in `ggplot2`.

Problem 2 (7 p)

a) Create a function you call `hilbert(n,m)` that creates a Hilbert matrix. The element (i, j) of the Hilbert matrix is defined as follows

$$H_{ij} = \frac{1}{i + j - 1}$$

where i is the row and j is the column in the matrix. The size of the matrix should be `n` $\times$ `m`.

```
hilbert(1,4)

     [,1] [,2]     [,3] [,4]
[1,]    1  0.5 0.333333 0.25


hilbert(2,2)

     [,1]      [,2]
[1,]  1.0 0.500000
[2,]  0.5 0.333333
```

b) Calculate $\det\left(\mathbf{H}^T\mathbf{H}\right)$ where $\mathbf{H}$ is a 5×5 Hilbert matrix.

c) Create a test suite for this function (using `testthat`) that checks that (1) the result of the function is a matrix and (2) that one of the examples above will be returned using the function.

Problem 3 (7 p)

a) Create a function you call `my_tidy_titanic_data()` that should return a tidy dataset from the Titanic dataset in R. You should use `tidyr` and `dplyr` in the function.

The `Titanic` dataset is a three dimensional table, below is how you should convert it to a `data.frame` (and then make it tidy).

```
# Load and convert the titanic dataset
data("Titanic")
x <- as.data.frame(as.matrix(ftable(Titanic)))
# The resulting function
head(my_tidy_titanic_data())

  class    sex   age survived counts
1   1st   Male Child       No      0
2   1st   Male Adult       No    118
3   1st Female Child       No      0
4   1st Female Adult       No      4
5   2nd   Male Child       No      0
6   2nd   Male Adult       No    154
```

b) Create a new function you call `aggregate_away_sex(x)` that takes a tidy titanic dataset as variable `x` and returns a dataset where the variable `sex` has been aggregated together in each group. See the example below. You should use `dplyr` functions.

```
tita<- my_tidy_titanic_data()
head(aggregate_away_sex(tita))

  class   age survived counts
1   1st Adult       No    122
2   1st Adult      Yes    197
3   1st Child       No      0
4   1st Child      Yes      6
5   2nd Adult       No    167
6   2nd Adult      Yes     94
```

Good luck!