

# Homework on optimization algorithms.

## P8160 Advanced Statistical Computing

### Problem 1:

Design an optimization algorithm to find the minimum of the continuously differentiable function

$$f(x) = -e^{-x} \sin(x)$$

on the closed interval  $[0, 1.5]$ . Write out your algorithm and implement it into **R**.

**Answer: your answer starts here...**

```
#R codes:
```

### Problem 2:

The Poisson distribution, written as

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

for  $\lambda > 0$ , is often used to model “count” data — e.g., the number of events in a given time period.

A Poisson regression model states that

$$Y_i \sim \text{Poisson}(\lambda_i),$$

where

$$\log \lambda_i = \alpha + \beta x_i$$

for some explanatory variable  $x_i$ . The question is how to estimate  $\alpha$  and  $\beta$  given a set of independent data  $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$ .

1. Generate a random sample  $(x_i, Y_i)$  with  $n = 500$  from the Poisson regression model above. You can choose the true parameters  $(\alpha, \beta)$  and the distribution of  $X$ .
2. Write out the likelihood of your simulated data, and its Gradient and Hessian functions.
3. Develop a modified Newton-Raphson algorithm that allows the step-halving and re-direction steps to ensure ascent directions and monotone-increasing properties.
4. Write down your algorithm and implement it in R to estimate  $\alpha$  and  $\beta$  from your simulated data.

**Answer: your answer starts here...**

```
#R codes:
```

### Problem 3:

The data *breast-cancer.csv* have 569 row and 33 columns. The first column **ID** labels individual breast tissue images; The second column **Diagnosis** identifies if the image is coming from cancer tissue or benign cases (M=malignant, B = benign). There are 357 benign and 212 malignant cases. The other 30 columns correspond to mean, standard deviation and the largest values (points on the tails) of the distributions of the following 10 features computed for the cellnuclei;

- radius (mean of distances from center to points on the perimeter)
- texture (standard deviation of gray-scale values)
- perimeter
- area
- smoothness (local variation in radius lengths)
- compactness ( $\text{perimeter}^2 / \text{area} - 1.0$ )
- concavity (severity of concave portions of the contour)
- concave points (number of concave portions of the contour)
- symmetry
- fractal dimension ("coastline approximation" - 1)

The goal is to build a predictive model based on logistic regression to facilitate cancer diagnosis;

1. Build a logistic model to classify the images into malignant/benign, and write down your likelihood function, its gradient and Hessian matrix.
2. Build a logistic-LASSO model to select features, and implement a path-wise coordinate-wise optimization algorithm to obtain a path of solutions with a sequence of descending  $\lambda$ 's.
3. Write a report to summarize your findings.