

Multivariate Optimization 4

Exercise 1: Newton-Raphson and Gauss-Newton

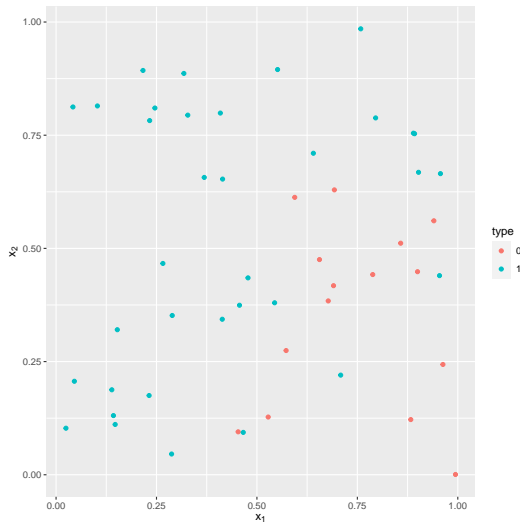
You are given the following data situation:

```
library(ggplot2)

set.seed(123)

# simulate 50 binary observations with noisy linear decision boundary
n = 50
X = matrix(runif(2*n), ncol = 2)
X_model = cbind(1, X)
y = -((X_model %*% c(0.3, -1, 1) + rnorm(n, 0, 0.3) < 0) - 1)
df = as.data.frame(X)
df$type = as.character(y)

ggplot(df) +
  geom_point(aes(x = V1, y = V2, color=type)) +
  xlab(expression(x[1])) +
  ylab(expression(x[2]))
```



In the following we want to estimate a model $\pi : \mathbb{R}^2 \rightarrow [0, 1], (x_1, x_2) \mapsto \frac{1}{1 + \exp((1, x_1, x_2)^\top \theta)}$ such that it minimizes the Brier-loss, i.e., $\mathcal{R}_{\text{emp}} = \sum_{i=1}^n \|y^{(i)} - \pi(\mathbf{x}^{(i)})\|_2^2$.

(a) Show that the gradient

$$\nabla_{\theta} \mathcal{R}_{\text{emp}} = \sum_{i=1}^n 2 \frac{y^{(i)} (\exp(f^{(i)})) - (\exp(-f^{(i)}) + 1)^{-1}}{(\exp(f^{(i)}) + 1)^2} \tilde{\mathbf{x}}^{(i)}$$

where $\tilde{\mathbf{x}}^{(i)} = (1, x_1^{(i)}, x_2^{(i)})^\top$ and $f^{(i)} = \tilde{\mathbf{x}}^{(i)\top} \theta$

(b) Show that the Hessian $\nabla_{\theta}^2 \mathcal{R}_{\text{emp}} = \sum_{i=1}^n 2 \frac{\exp(f^{(i)}) (y^{(i)} (-\exp(2f^{(i)}) + 1) - 1 + 2 \exp(f^{(i)}))}{(\exp(f^{(i)}) + 1)^4} \tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)\top}$

(c) Show that \mathcal{R}_{emp} is not convex in general

- (d) Write an R script to find an optimal model via Newton-Raphson (do 30 iterations, $\mathbf{x}^{[0]} = \mathbf{0}$).
- (e) Explain why Gauss-Newton is applicable here and write an R script to find an optimal model via Gauss-Newton (do 30 iterations, $\mathbf{x}^{[0]} = \mathbf{0}$).