

## Multivariate Optimization 1

### Exercise 1: Gradient Descent

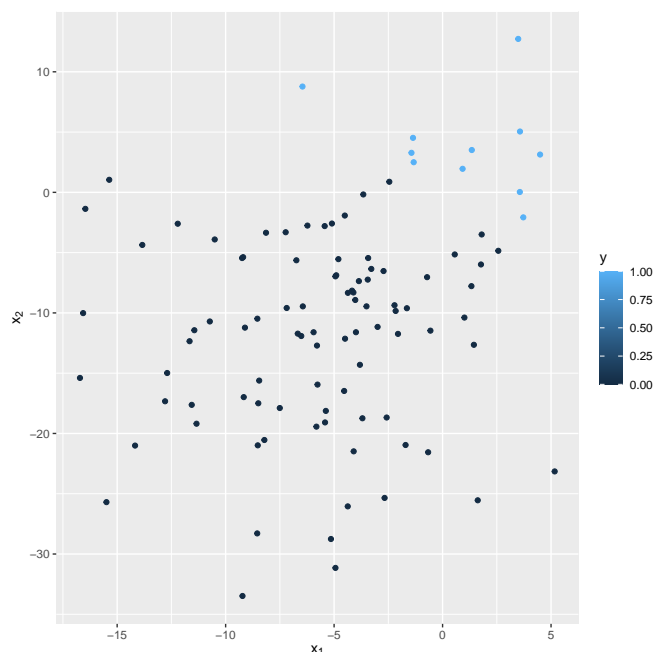
You are given the following data situation:

```
library(ggplot2)

set.seed(314)
n <- 100
X = cbind(rnorm(n, -5, 5),
          rnorm(n, -10, 10))
X_design = cbind(1, X)

z <- 2*X[,1] + 3*X[,2]
pr <- 1/(1+exp(-z))
y <- as.integer(pr > 0.5)
df <- data.frame(X = X, y = y)

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



In the following we want to estimate a logistic regression without intercept via gradient descent<sup>1</sup>.

- (a) The data situation is called complete separation, i.e., the classes can be perfectly classified with a linear classifier. Show that in this situation if  $\tilde{\theta}$  perfectly classifies the data then:  
 $\mathcal{R}_{\text{emp}}(\tilde{\theta}) > \mathcal{R}_{\text{emp}}(\alpha \tilde{\theta})$  with  $\alpha > 1$ .
- (b) Visualize  $\mathcal{R}_{\text{emp}}$  in  $[-1, 4] \times [-1, 4]$ .

<sup>1</sup>We chose this algorithm for educational purposes; in practice, we typically use second order algorithms.

- (c) Find the gradient of  $\mathcal{R}_{\text{emp}}$  for arbitrary  $\boldsymbol{\theta}$ .
- (d) Solve the logistic regression via gradient descent. Use a step size  $\alpha = 0.01$ , starting point  $\boldsymbol{\theta}^{[0]} = (0, 0)^\top$  and train for 500 steps. Repeat this with  $\alpha = 0.02$ . Explain your observation.  
*Hint:* a)
- (e) Repeat d) but add an L2 penalization term (with  $\lambda = 1$ ) to the objective. What do you observe now?
- (f) Visualize the regularized  $\mathcal{R}_{\text{emp}}$  in  $[-1, 4] \times [-1, 4]$ .
- (g) Repeat e) but with backtracking. Set  $\gamma = 0.9$  and  $\tau = 0.5$