

Univariate Optimization 1

Exercise 1: Golden Ratio, Brent's Method

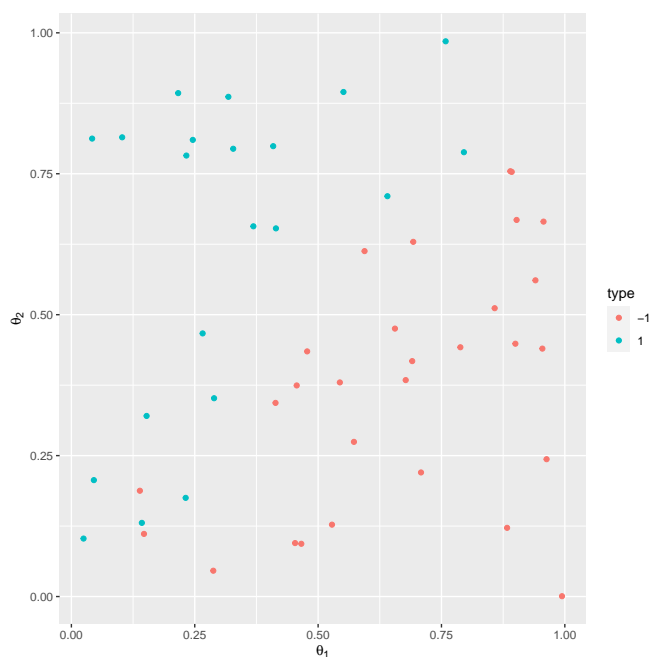
You are given the following data situation:

```
library(ggplot2)

set.seed(123)

X = matrix(runif(100), ncol = 2)
y = -((X %>% c(-1, 1) + rnorm(50, 0, 0.1) < 0) * 2 - 1)
df = as.data.frame(X)
df$type = as.character(y)

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



In the following we want to estimate a linear SVM without intercept and with $\lambda = 1$. We assume we know that $\theta_2 = 2$.

- Show that if $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is convex then $g_c : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x, c) \quad \forall c \in \mathbb{R}$ is convex.
- Explain why the non-geometric primal linear SVM formulation should be used rather than the geometric one if we want to find θ_1 via the golden ratio algorithm¹.
- Find θ_1 via the golden ratio algorithm. Implement the algorithm in R. For the termination criterion, use an absolute error of 0.01. Use $[-3, 3]$ as the starting interval.

¹We choose this algorithm for educational purposes; in practice, we typically use more advanced algorithms.

- (d) Given the three points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ show that the parameters $a, b, c \in \mathbb{R}$ of the interpolating parabola can be found via $\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$ when the parabola equation is given by $p(x) = ax^2 + bx + c$.
- (e) Find θ_1 via Brent's method². Implement a simplified version³ of the algorithm in R. For the termination criterion, use an absolute error of 0.01. Use $[-3, 3]$ as the starting interval. For the first step, use a golden ratio step.
- (f) Now, assume we do not know θ_2 . Our initial guess is $\theta_2 = 0$. We now alternately minimize w.r.t. either θ_1 or θ_2 via the golden ratio method (the starting interval is always reset to $[-3, 3]$) while the other parameter is held constant. We switch to minimizing the other parameter when the absolute error is smaller than 0.01. Repeat this procedure 10 times.
- (g) How does the optimization trace of f) look in parameter space?
- (h) Can we find the optimal parameters of the linear SVM with our procedure in f)?
Hint: a)

²We choose this algorithm for educational purposes; in practice, we typically use more advanced algorithms.

³Only check if the proposed point is in the current interval