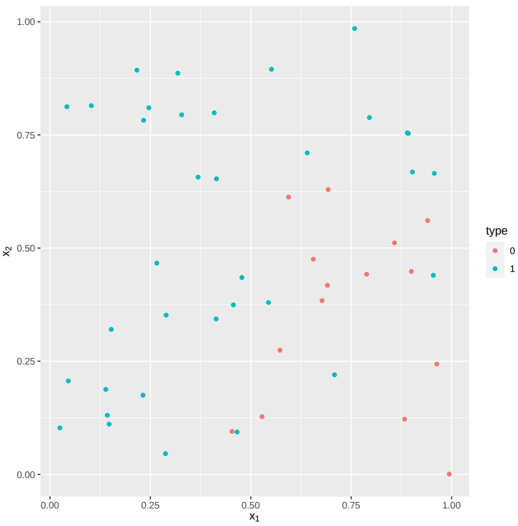
Multivariate Optimization 4

Solution 1: Newton-Raphson and Gauss-Newton

You are given the following data situation:

```r
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]))
```



(a) Define $s : \mathbb{R} \to \mathbb{R}, f \mapsto \frac{1}{1+\exp(f)}$.

$\nabla_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}} = \nabla_{\boldsymbol{\theta}} \sum_{i=1}^n \|y^{(i)} - f(\mathbf{x}^{(i)})\|_2^2 = \sum_{i=1}^n \frac{d}{df} \|y^{(i)} - s(f^{(i)})\|_2^2 \cdot \nabla_{\boldsymbol{\theta}} f^{(i)}$

$= \sum_{i=1}^n 2 \frac{y^{(i)}(\exp(f^{(i)})+1)-1}{\exp(f^{(i)})+1} \cdot \frac{\exp(f^{(i)})}{(\exp(f^{(i)})+1)^2} \tilde{\mathbf{x}}^{(i)}$

$= \sum_{i=1}^n 2 \frac{y^{(i)}(\exp(f^{(i)})+1)-1}{\exp(f^{(i)})+1} \cdot \frac{\frac{\exp(f^{(i)})}{\exp(f^{(i)})+1}}{\exp(f^{(i)})+1} \tilde{\mathbf{x}}^{(i)}$

$= \sum_{i=1}^n 2 \frac{y^{(i)}(\exp(f^{(i)})) - \frac{\exp(f^{(i)})}{\exp(f^{(i)})+1}}{(\exp(f^{(i)})+1)^2} \tilde{\mathbf{x}}^{(i)}$

$= \sum_{i=1}^n 2 \frac{y^{(i)}(\exp(f^{(i)})) - (\exp(-f^{(i)})+1)^{-1}}{(\exp(f^{(i)})+1)^2} \tilde{\mathbf{x}}^{(i)}$

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

$= \sum_{i=1}^n 2 \frac{(y^{(i)}(\exp(f^{(i)})) - (\exp(-f^{(i)})+1)^{-2} \exp(-f^{(i)}))(\exp(f^{(i)})+1)^2 - (y^{(i)}(\exp(f^{(i)})) - (\exp(-f^{(i)})+1)^{-1}) \cdot 2(\exp(f^{(i)})+1)\exp(f^{(i)})}{(\exp(f^{(i)})+1)^4} \tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)\top}$

$$
\begin{aligned}
&= \sum_{i=1}^{n} 2\frac{y^{(i)}\exp(f^{(i)})(\exp(f^{(i)})+1)^2 - \exp(f^{(i)}) - (2y^{(i)}\exp(f^{(i)})(\exp(f^{(i)})+1)+2\exp(f^{(i)}))\exp(f^{(i)})}{(\exp(f^{(i)})+1)^4}\tilde{\mathbf{x}}^{(i)}\tilde{\mathbf{x}}^{(i)\top} \\
&= \sum_{i=1}^{n} 2\frac{\exp(f^{(i)})(y^{(i)}(\exp(f^{(i)})+1)^2 - 1 - 2y^{(i)}\exp(f^{(i)})(\exp(f^{(i)})+1)+2\exp(f^{(i)}))}{(\exp(f^{(i)})+1)^4}\tilde{\mathbf{x}}^{(i)}\tilde{\mathbf{x}}^{(i)\top} \\
&= \sum_{i=1}^{n} 2\frac{\exp(f^{(i)})(y^{(i)}(\exp(2f^{(i)})+2\exp(f^{(i)})+1-2\exp(2f^{(i)})-2\exp(f^{(i)}))-1+2\exp(f^{(i)}))}{(\exp(f^{(i)})+1)^4}\tilde{\mathbf{x}}^{(i)}\tilde{\mathbf{x}}^{(i)\top} \\
&= \sum_{i=1}^{n} 2\frac{\exp(f^{(i)})(y^{(i)}(-\exp(2f^{(i)})+1)-1+2\exp(f^{(i)}))}{(\exp(f^{(i)})+1)^4}\tilde{\mathbf{x}}^{(i)}\tilde{\mathbf{x}}^{(i)\top}
\end{aligned}
$$

(c) Assume, e.g., there is only one observation with $y^{(1)} = 0$ then

$$
\nabla^2_{\boldsymbol{\theta}}\mathcal{R}_{\text{emp}} = \frac{2\exp(f^{(1)})(2\exp(f^{(1)})-1)}{(\exp(f^{(1)})+1)^4}\underbrace{\tilde{\mathbf{x}}^{(1)}\tilde{\mathbf{x}}^{(1)\top}}_{\text{p.s.d.}}.
$$

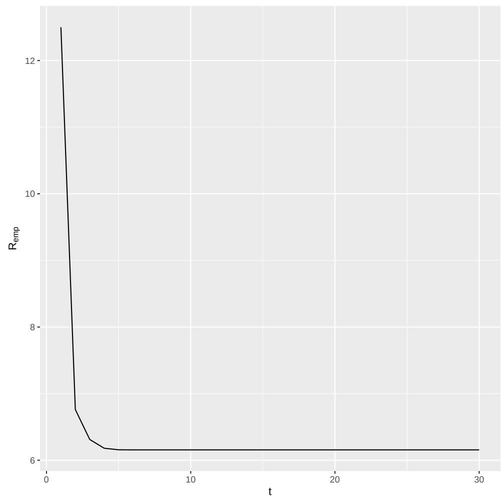
If a p.s.d. matrix is multiplied with a negative number it becomes a n.s.d. matrix, i.e., $\nabla^2_{\boldsymbol{\theta}}\mathcal{R}_{\text{emp}}$ is n.s.d. if $2\exp(f^{(1)}) < 1 \iff f^{(i)} < \ln(0.5)$. This condition trivially holds, e.g., if $\boldsymbol{\theta} = (\ln(0.5) - 1, 0, 0)^\top$.

(d) For Newton-Raphson, we need to solve in each update step

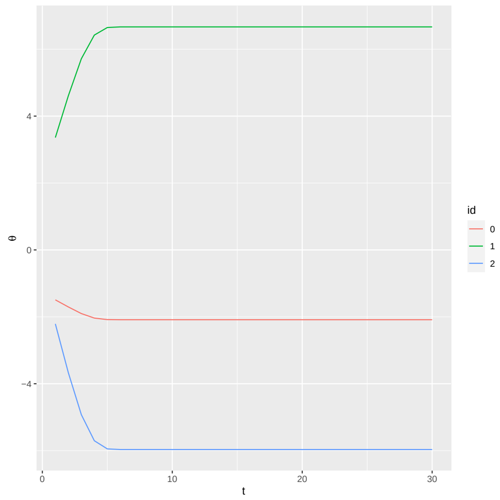
$$
\nabla^2_{\boldsymbol{\theta}}\mathcal{R}_{\text{emp}}\mathbf{d} = -\nabla_{\boldsymbol{\theta}}\mathcal{R}_{\text{emp}}
$$

for the descend direction $\mathbf{d}$.

```r
theta = c(0, 0, 0)
remps = NULL
thetas = NULL

for(i in 1:30){
  exp_f = exp(X_model %*% theta)
  remps = rbind(remps, sum((y - 1/(1+exp_f))^2))

  hess = t(X_model) %*%
    (c((2 * exp_f*(2 * exp_f - y*(exp_f^2 - 1) - 1))/(exp_f + 1)^4 ) * X_model)
  grad = c(t(2*(y * exp_f - (1 + exp_f^-1)^-1) / (exp_f + 1)^2) %*% X_model)
  theta = theta + solve(hess, -grad)

  thetas = rbind(thetas, theta)
}

ggplot(data.frame(remps, t=1:nrow(remps)), aes(x=t, y=remps)) +
  geom_line() + ylab(expression(R[emp]))
```



```r
ggplot(data.frame(theta = c(thetas), t=rep(1:nrow(thetas),3),
                  id = as.factor(rep(c(0, 1, 2), each= nrow(thetas)))),
       aes(x = t, y=theta)) +
  geom_line(aes(color = id)) + ylab(expression(theta))
```




```r
theta

## [1] -2.087122  6.667438 -5.967500
```

(e) In this case, we can apply Gauss-Newton since $\mathcal{R}_{\text{emp}}$ is the squared sum of the residuals

$$\mathbf{r} = (y^{(1)} - \pi(\mathbf{x}^{(1)}), \dots, y^{(n)} - \pi(\mathbf{x}^{(n)}))^\top.$$

Here, for the update step we need to compute

$$\nabla_{\boldsymbol{\theta}} \mathbf{r} = \begin{pmatrix} \frac{\exp(f^{(1)})}{(1+\exp(f^{(1)}))^2} \tilde{\mathbf{x}}^{(1)\top} \\ \vdots \\ \frac{\exp(f^{(n)})}{(1+\exp(f^{(n)}))^2} \tilde{\mathbf{x}}^{(n)\top} \end{pmatrix}$$

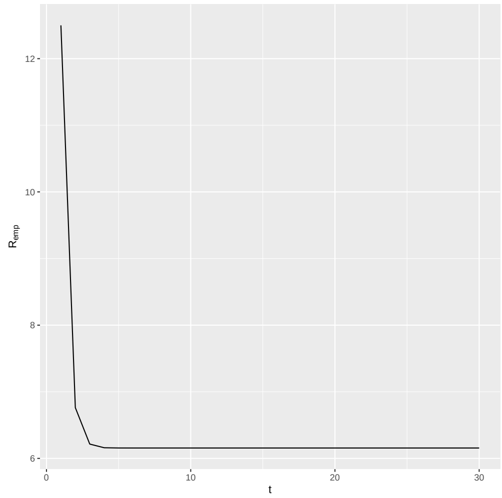
For Gauss-Newton, we solve in each update step

$$(\nabla_{\boldsymbol{\theta}} \mathbf{r}^\top \nabla_{\boldsymbol{\theta}} \mathbf{r}) \cdot \mathbf{d} = -\nabla_{\boldsymbol{\theta}} \mathbf{r}^\top \cdot \mathbf{r}$$

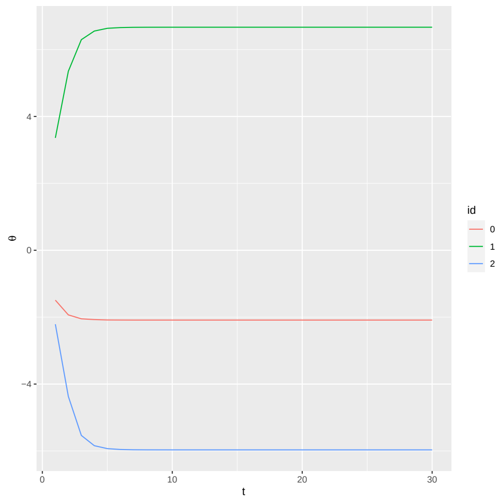
for the descend direction $\mathbf{d}$.

```r
theta = c(0, 0, 0)
remps = NULL
thetas = NULL

for(i in 1:30){
  exp_f = exp(X_model %*% theta)
  remps = rbind(remps, sum((y - 1/(1+exp_f))^2))

  res = (y-1/(1+exp_f))
  grad_res = c(exp_f / (exp_f + 1)^2) * X_model

  theta = c(theta + solve(t(grad_res) %*% grad_res, -t(grad_res) %*% res))
  thetas = rbind(thetas, theta)
}
```

```
ggplot(data.frame(remps, t=1:nrow(remps)), aes(x=t, y=remps)) +
  geom_line() + ylab(expression(R[emp]))
```




```
ggplot(data.frame(theta = c(thetas), t=rep(1:nrow(thetas),3),
                  id = as.factor(rep(c(0, 1, 2), each= nrow(thetas)))),
       aes(x = t, y=theta)) +
  geom_line(aes(color = id)) + ylab(expression(theta))
```




```
theta

## [1] -2.087122  6.667438 -5.967500
```