STA212

Anthony Kalaydjian - Mathieu Occhipinti

2023-04-29

```
rm(list=ls())
setwd(getwd())
library(ggplot2)
set.seed(150)
```

Exercice 1 : Modélisation probabiliste

(a)

(b)

Le résultat précédent nous indique que pour toute règle de classification empirique, $\hat{g_n}$ issue de l'estimateur $\hat{\eta}$ de η, son risque associé est borné par le risque minimal issu de la règle de Bayes $R(g^*)$ auquel on ajoute un terme d'erreur d'estimation de η.

Exercice 2 : Classification multi-classes

(a)

(b)

(c)

Exercice 3 : Implémentation d'un perceptron (origine des SVM)

(a)

Supposons que l'on soit à l'itération t de l'algorithme, et tel que $m(\theta^t) \neq \emptyset$.\ Soit alors $i \in m(\theta^t)$

D'une part:

$$\begin{aligned}
\langle \theta^{t+1}, \theta^* \rangle &= \langle \theta^t + y_i x_i, \theta^* \rangle \\
&= \langle \theta^t, \theta^* \rangle + y_i \langle x_i, \theta^* \rangle \\
&\geq \langle \theta^t, \theta^* \rangle + ||\theta^*||_2 \delta
\end{aligned}$$

Par Cauchy-Schwarz:

$$
\begin{aligned}
||\theta^{t+1}||_2||\theta^*||_2 \geq \langle \theta^{t+1}, \theta^* \rangle &\geq \langle \theta^t, \theta^* \rangle + ||\theta^*||_2\delta \\
&\geq \langle \theta^{t-1}, \theta^* \rangle + 2||\theta^*||_2\delta \\
&\geq \ldots \\
&\geq \cancel{\langle 0, \theta^* \rangle} + t||\theta^*||_2\delta \\
||\theta^{t+1}||_2||\theta^*||_2 &\geq t||\theta^*||_2\delta
\end{aligned}
$$

Donc:

$$\boxed{||\theta^{t+1}||_2 \geq t\delta} \tag{1}$$

D'autre part:

$$
\begin{aligned}
||\theta^{t+1}||^2 &= ||\theta^t + y_i x_i||^2 \\
&= ||\theta^t||^2 + 2y_i\langle \theta^t, x_i \rangle + ||y_i x_i||^2 \\
&\leq ||\theta^t||^2 + ||x_i||^2 \\
&\leq ||\theta^t||^2 + R^2 \\
&\leq ||\theta^{t-1}||^2 + 2R^2 \\
&\leq \ldots \\
&\leq \cancel{||\theta^0||^2} + tR^2 \\
||\theta^{t+1}||^2 &\leq tR^2
\end{aligned}
$$

Donc:

$$\boxed{||\theta^{t+1}||^2 \leq tR^2} \tag{2}$$

Finalement, d'après **??** et **??**, on a :

$$t^2\delta^2 \leq ||\theta^{t+1}||^2 \leq tR^2$$

Donc $\boxed{t \leq \dfrac{R^2}{\delta^2}}$

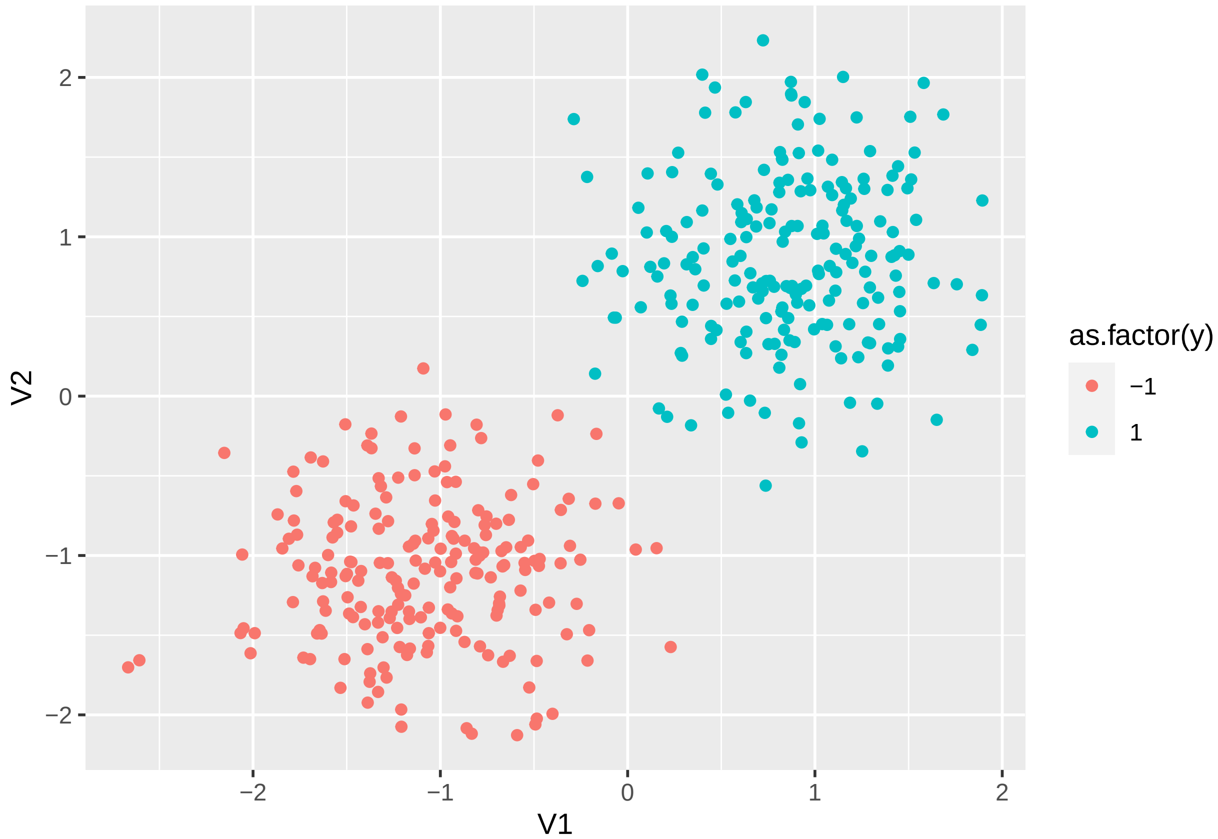
Ainsi, on a montré que si $m(\theta^t) \neq \emptyset$, alors $t \leq \frac{R^2}{\delta^2}$

Donc, au delà de $T = \frac{R^2}{\delta^2}$ itérations, $m(\theta^T)$ sera vide et l'algorithme aura convergé.

(b)

Importation des données :

```
load(file="X_y.rda")
df <- as.data.frame(cbind(X, y))
names(df) <- c("V1", "V2", "V3", "y")
plt1 <- ggplot(data=df) + aes(x=V1, y=V2, z=y, color=as.factor(y)) + geom_point()
plt1
```



La variable V3 est une variable d'intercept.

Algorithme perceptron

```
perceptron <- function(X, y){
  theta <- c(0, 0, 0)
  n <- nrow(X)
  m <- seq(1, n)
  counter <- 0

  while (length(m) != 0){
    #sample a random item from m
    index = sample(m, 1)

    #update theta
    theta <- theta + y[index]*X[index,]

    #calculate the new m
    temp <- sapply(X=seq(1, n), FUN=function(k) theta%*%X[k,])
    criterion <- y*temp
    m <- which(criterion<0)
    counter <- counter + 1
  }
  return(list(theta=theta, count=counter))
}

res <- perceptron(X, y)
```

```r
theta.star <- res$theta
count.star <- res$count

theta.star
```

```
## [1] 3.438710 4.537851 1.000000
```

```r
count.star
```

```
## [1] 5
```

L'algorithme converge en 5 itérations, et nous trouve la valeur de $\theta^* = (3.438710, 4.537851, 1.000000)^T$.

plot

```r
plt1 + geom_abline(intercept=-theta.star[3]/theta.star[2], slope=-theta.star[1]/theta.star[2])
```