TP08 - EM Algorithm

Machine Learning

Carranza-Alarcón Yonatan-Carlos[1]

[1]UMR CNRS 7253 Heudiasyc
Université de technologie de Compiègne

November 25, 2020

EM Algorithm

Soit la densité du mélange suivante

$$p(x;\theta) = \pi\varphi(x;\mu,\sigma) + (1-\pi)\phi(x), \tag{1}$$

où $\varphi(\cdot;\mu,\sigma)$ est la fonction de densité de la loi normale $\mathcal{N}(\mu,\sigma)$ et $\phi(x)$ est la fonction de densité de la loi l'uniforme $\mathcal{U}(-a,a)$(donc $\forall x, \phi(x) = \frac{1}{2a}$).

Problème: Soit l'échantillon $X = (X_1,\ldots,X_n)$ issu de la densité $p(x,\theta)$ où $\theta = (\mu,\sigma,\pi)$ est inconnu. Donc, comment peut-on estimer les paramètres θ?

1. Bayesian inference (MCMC/ABC)
2. Frequentist inference (Maximum likelihood estimation)

EM Algorithm

Maximum de vraisemblance: $\log L(\theta; x_1, ..., x_n)$

$$\ell(\theta, X) = \sum_{i=1}^{n} \log\left(\pi \frac{1}{\sigma\sqrt{2(3.14)}} exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right) + (1-\pi)\frac{1}{2a}\right)$$

Trop difficile, même parfois impossible, à résoudre!!!

Première dérivé $d(\mu)\ell(\theta, X) = 0$ et dérivé seconde $d(\mu)^2\ell(\theta, X) < 0$

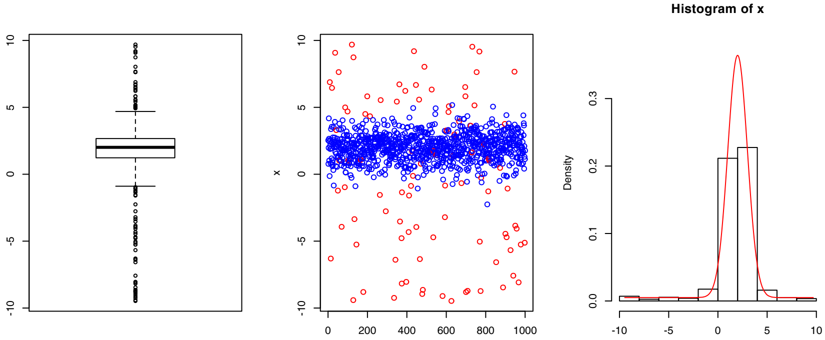
$$\sum_{i=1}^{n} \frac{\pi \frac{1}{\sigma\sqrt{2(3.14)}} exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)}{\pi \frac{1}{\sigma\sqrt{2(3.14)}} exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right) + (1-\pi)\frac{1}{2a}} \left(-\frac{(x_i-\mu)}{\sigma^2}\right)(-1) = 0 \quad (2)$$

On verra que le terme en rouge est égal à l'espérance de notre variable latente, i.e. $\mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i]$

Génération des données de la densité $p(x;\theta)$

$$\pi = 0.9, \mu = 2, \sigma = 1, a = 10, n = 1000$$

```r
x <- rep(NA, n)
y <- rep(NA, n)
for(i in 1:n){
  y[i] <- sample(c(1,0), size=1, prob=c(pi, 1-pi))
  if(y[i]==1){
    x[i] <- rnorm(1,mean=mu,sd=sigma)
  }else{
    x[i] <- runif(1,min=-a,max=a)
  }
}
```



EM Algorithm - Formallement

On suppose qu'il exist variable aléatoire latente $Y \in \{1, \dots, k\}$ telle que (X, Y) peut partitioner le nuage de point X en k catégories.

E-Step:

$$
\begin{aligned}
Q(\theta, \theta^{(t)}) &= \mathbb{E}_{Y|X,\theta^{(t)}} [\log L(\theta; X, Y)] && (3)\\
&= \sum_{Y} P(Y|X, \theta^{(t)}) \log(L(\theta; X, Y)) && (4)
\end{aligned}
$$

M-Step:

$$
\theta^{(t+1)} = \arg\max_{\theta} Q(\theta, \theta^{(t)}) \quad (5)
$$

Loop: $|\ell(\theta^{(t)}; X) - \ell(\theta^{(t+1)}; X)| < \epsilon$

EM Algorithm - officieusement

E-Step: $\forall i \in \{1, \ldots, n\}$ et $\theta^{(t)} = (\mu^{(t)}, \sigma^{(t)}, \pi^{(t)})$ fixé.

$$\hat{y}_i^{(t)} = \mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i] \tag{6}$$

M-Step:

$$\theta^{(t+1)} = \arg\max_{\theta} \mathbb{E}_{Y|X,\theta^{(t)}} [\log L(\theta; X, Y)] \tag{7}$$

Un cas particulier: Quand $y_i \in \{0, 1\}$

$$\hat{y}_i^{(t)} = P(y_i = 1|x_i, \theta^{(t)}) = \frac{P(y_i = 1, x_i|\theta^{(t)})}{P(x_i|\theta^{(t)})} \tag{8}$$

$$= \frac{\varphi(x; \mu^{(t)}, \sigma^{(t)})\pi^{(t)}}{\varphi(x; \mu^{(t)}, \sigma^{(t)})\pi^{(t)} + (1 + pi^{(t)})1/2a} \tag{9}$$

EM Algorithm

Comme, on procède (pour plus de détails EM derivation.pdf (moodle))

$$\mathbb{E}_{Y|X,\theta^{(t)}}\left[\log L(\theta; X, Y)\right] = \mathbb{E}_{Y|X,\theta^{(t)}}\left[\sum_{i=1}^{n}\log p(x_i, y_i|\theta)\right] = \sum_{i=1}^{n}\mathbb{E}_{y_i|x_i,\theta^{(t)}}\left[\log p(x_i, y_i|\theta)\right]$$

On calcule $\log p(x_i, y_i|\theta)$:

$$\log p(x_i, y_i|\theta) = y_i \log(\pi\phi(x_i, \mu, \sigma)) + (1 - y_i)\log((1 - pi)\frac{1}{2a})$$

Et finalement l'espérance conditionnelle $\mathbb{E}_{y_i|x_i,\theta^{(t)}}\left[\log p(x_i, y_i|\theta)\right]$

$$\mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i]\log(\pi\phi(x_i, \mu, \sigma)) + (1 - \mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i])\log((1 - \pi)\frac{1}{2a})$$

$\mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i]$ impossible à calculer sans $\theta^{(t)}$, et nécessaire pour maximiser l'espérance $\mathbb{E}_{Y|X,\theta^{(t)}}\left[\log L(\theta; X, Y)\right]$ dans l'étape M-step.

EM Algorithm (E-Step) in R

Pour un $\theta^{(t)} = (\mu^{(t)}, \sigma^{(t)}, \pi^{(t)})$

$$\hat{y}_i = \mathbb{E}_{y_i|x_i,\theta^{(t)}}[y_i] = \frac{\varphi(x;\mu^{(t)},\sigma^{(t)})\pi^{(t)}}{\varphi(x;\mu^{(t)},\sigma^{(t)})\pi^{(t)} + (1-\pi^{(t)})1/2a}$$

```
# E-step
phi <- dnorm(x,mean=thetat[1],sd=thetat[2])
y<- phi*thetat[3]/(phi*thetat[3]+1/2a*(1-thetat[3]))
```

EM Algorithm (M-Step) in R

M-Step: En dérivant l'espérance $Q(\theta, \theta^{(t)})$, on obtient:

$$\hat{\pi}^{(t+1)} = \sum_{i=1}^{n} \hat{y}_i / n \tag{10}$$

$$\hat{\mu}^{(t+1)} = \frac{\sum_{i=1}^{n} \hat{y}_i x_i}{\sum_{i=1}^{n} \hat{y}_i} \tag{11}$$

$$\hat{\sigma}^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{n} \hat{y}_i (x_i - \mu^{(t+1)})^2}{\sum_{i=1}^{n} \hat{y}_i}} \tag{12}$$

$$\tag{13}$$

Donc le nouveau $\theta^{(t+1)} = (\hat{\mu}^{(t+1)}, \hat{\sigma}^{(t+1)}, \hat{\pi}^{(t+1)})$

```r
# M-step
S <- sum(y)
pi <- S/n
mu <- sum(x*y)/S
sig <- sqrt(sum(y*(x-mu)^2)/S)
theta <- c(mu, sig, pi)
```

EM Algorithm (Condition d'arrêt) in R

Loop: $|\ell(\theta^{(t)}; X) - \ell(\theta^{(t+1)}; X)| < \epsilon$‘

$$\ell(\theta_*; X) = \sum_{i=1}^{n} \log \left(\pi_* \frac{1}{\sigma_* \sqrt{2(3.14)}} exp \left(-\frac{1}{2} \frac{(x_i - \mu_*)^2}{\sigma_*^2} \right) + (1 - \pi_*) \frac{1}{2a} \right)$$

```r
loglik<- function(theta ,x){
  phi <- dnorm(x,mean=theta[1],sd=theta[2])
  logL <- sum(log(theta[3]*phi+(1-theta[3])*1/2a))
  return(logL)
}
```

EM Algorithm in R

```r
logL0<- loglik(thetat,x)
while(go_on){
 t <- t+1
 # E-step
 phi <- dnorm(x,mean=thetat[1],sd=thetat[2])
 y<- phi*thetat[3]/(phi*thetat[3]+c*(1-thetat[3]))
 # M-step
 S <- sum(y)
 pi <- S/n
 mu <- sum(x*y)/S
 sig <- sqrt(sum(y*(x-mu)^2)/S)
 theta <- c(mu,sig,pi)
 logL<-loglik(theta,x)
 if (logL-logL0 < epsi){
  go_on <- FALSE
 }
    logL0 <- logL
    theta0<-theta print(c(t,logL))
}
```