Mélange de gaussiens EM

ZIDAN Loubna

2023-02-15

```
library(mvtnorm)
library(graphics)
library(ggplot2)
```

Introduction

Expectation-Maximization (EM) est un algorithme itératif pour trouver des estimations de probabilité maximale de paramètres dans des modèles statistiques, où le modèle dépend de variables latentes non observées.

L'itération EM alterne entre l'exécution d'une étape (E), qui crée une fonction pour l'espérance de la log-vraisemblance évaluée à l'aide de l'estimation actuelle des paramètres.

Et une étape de maximisation (M), qui calcule les paramètres maximisant la log-vraisemblance attendue. vraisemblance trouvée à l'étape E.

Répéter ces étapes jusqu'à ce que la convergence soit détectée.

Au cours de ce projet, nous allons implémenter l'algorithme EM pour la distribution du mélange gaussien, en utilisant une variable latente Z.

Plan

1-Implementer la fonction Estep qui retourne les probabilités d'appartenance de chaque point à chaque cluster

2-Implementer la fonction Mstep qui retourne les parametres optimales du modele

3-Implementer la fonction vraisemblance

4-Regrouper toutes ces fonctions en une seule function EM

La fonction EM va retourner:

-Les itérations avec les valeurs ve vraisemblances coresspondantes

-Le plot de vraisemblance

-Le plot des erreurs qui montre la convergence

-Les parametres optimaux correspondants à chaque cluster: (densité(pi), moyenne(mu), et variance(sigma))

-Gamma qui sont les probabilités d'appartenance de chaque point à chaque cluster

-Le cluster attribué à chaque point

-Les données colorées par cluster (plot final)

Nous allons appliquer notre modele sur deux base de données:

-La premiere est une base de données simulée multivariée à l'aide de la fonction (mvnorm) de deux dimensions avec 350 points (N).

Ses parametres seront générées en fonction de K (nombre de clusters) et degree(nombre de dimensions). Nous allons fixer k le nombre de clusters à 2.

-La deuxieme base de données sera Iris, nous allons sélectionner également 2 colonnes 3 et 4, et le nombre de clusters sera égale à 2 également.

L'algorithme se repete jusqu'à n.trial fois qui est égale à N, la condition d'arrete sera :

si la proportion ((nouvelle valeur de vraisemblance moins l'ancienne valeure obtenue)divisé par l'ancienne valeure) est inferieure à un épsilon e , on sort de la boucle!

Voici les fonctions Estep et Mstep:

on vas donc utiliser les fonctions E step et M Step pour trouver les paramètres de la distribution du mélange gaussien et les appliquer jusqu'à la convergence.

```
#Estep:

Estep <- function(data,pi,mu,sigma,K){

  result <- apply(data, 1, function(xt){

    gamma_i <- sapply(1:K,function(k) {

      pi[k] * dmvnorm(xt, mu[,k], sigma[,,k])
    })

    gamma_i / sum(gamma_i)#normalization
  })

 gamma<- t(result)
}
```

```
#Mstep:

Mstep <- function(gamma,data,K,N,degree) {
  gamma.sum <- apply(gamma,2,sum)
  new.pi <- gamma.sum/N;
  new.mu <- t(t(t(data) %*% gamma) / gamma.sum);
  new.sigma <- array(0, dim=c(degree, degree, K));


  for(k in 1:K) {
    sig <- matrix(0,degree,degree);
    for(n in 1:N) {
      sig <- sig + gamma[n, k] * (data[n,] %*% t(data[n,]));
    }
    new.sigma[,,k] <- sig / gamma.sum[k] - new.mu[,k] %*% t(new.mu[,k])
  }
  list(new.pi, new.mu, new.sigma);
}
```

Calcul de la vraisemblance marginale:

```
vraisemblance <- function(data, pi,mu,sigma,K){
```

```
  loglike = matrix(0, n.trial, K)
  for(k in 1:K) {
    loglike[,k] = pi[k] * dmvnorm(data, mu[,k], sigma[,,k])
    }
  loglike = sum(log(rowSums(loglike)))

return(sum(loglike))
}
```

Ensuite, nous allons regrouper tout dans une seule fonction EM comme suivant:

EM:

```
EM<-function(data,K,n.trial,e){

#initalisations aléatoires des parametres :

  pi <- runif(K)
  pi <- runif(K)/sum(pi)

  mu <- matrix(runif(K*degree), nrow=degree, ncol=K)

  sigma <- array(0, dim=c(degree, degree, K))
  for(k in 1:K){
    sigma[,,k] <- diag(runif(degree))
  }



#calculer l'error, la vraisemblance, et les parametres optimales:

  count <- 0
  errorlist    <-  rep(0,n.trial)
  logliklist   <-  rep(0,n.trial)

  for(i in 1:n.trial){
    count <- count +1
    old_loglik = vraisemblance(data,pi,mu,sigma,K)

    gamma <- Estep(data,pi,mu,sigma,K)
    result <- Mstep(gamma,data,K,N,degree)

    new.pi <- result[[1]]
    new.mu <- result[[2]]
    new.sigma <- result[[3]]

    error <-sum((new.pi-pi)^2) + sum((new.mu-mu)^2) + sum((new.sigma-sigma)^2)
    errorlist[i] <- error

    pi <- new.pi
    mu <- new.mu
    sigma <- new.sigma

    new_loglik = vraisemblance(data,pi,mu,sigma,K)
```

```
    logliklist[i+1] = new_loglik


    #critere d'arret:
    critere = abs((new_loglik - old_loglik)/old_loglik)
    if(critere < e) break

    cat("loglike à l\'étape ", i, " : ", new_loglik, '\n')
  }



  #Les resultats:

  #Afficher le plot de vraisemblance
  plot_lhood<-plot(logliklist[2:count],main='vraisemblance')

  #Afficher le plot de l'erreur
  plot_error<-plot(errorlist[2:count],log="y",main='error plot')


  #Afficher les parametres finales:
  for(k in 1:K) {
    cat('Cluster ', k, '\n')

    cat('Pi  : \n')
    print(pi[k])

    cat('mu : \n')
    print(mu[k])

    cat('Sigma : \n')
    print(sigma[k])
}


  #Afficher le cluster attribué à chaque point :
  clustering = apply(gamma, MARGIN = 1, which.max)
  plot_clustering<-plot(clustering, main='clustering plot')




  clustr<-data.frame(clustering)
  x<-data.frame(data[,1])
  y<-data.frame(data[,2])

  data_final<-data.frame(x,y,clustr)


  #Afficher les données colorées par cluster:
  data_plot<-ggplot(data_final) +
  aes(x = data...1., y = data...2., colour = clustering) +
```

```
  geom_point(shape = "circle", size = 2.45) +
  scale_color_gradient(low = "#4184C8", high = "#D65A0D") +
  theme_minimal() +
  theme(legend.position = "none")


  return(list(plot_lhood,plot_error,gamma,clustering,plot_clustering,data_plot))
}
```

Premiere expérience : Appliquer EM sur des données simulées

on utilise les paramètres ci-dessous pour générer les données

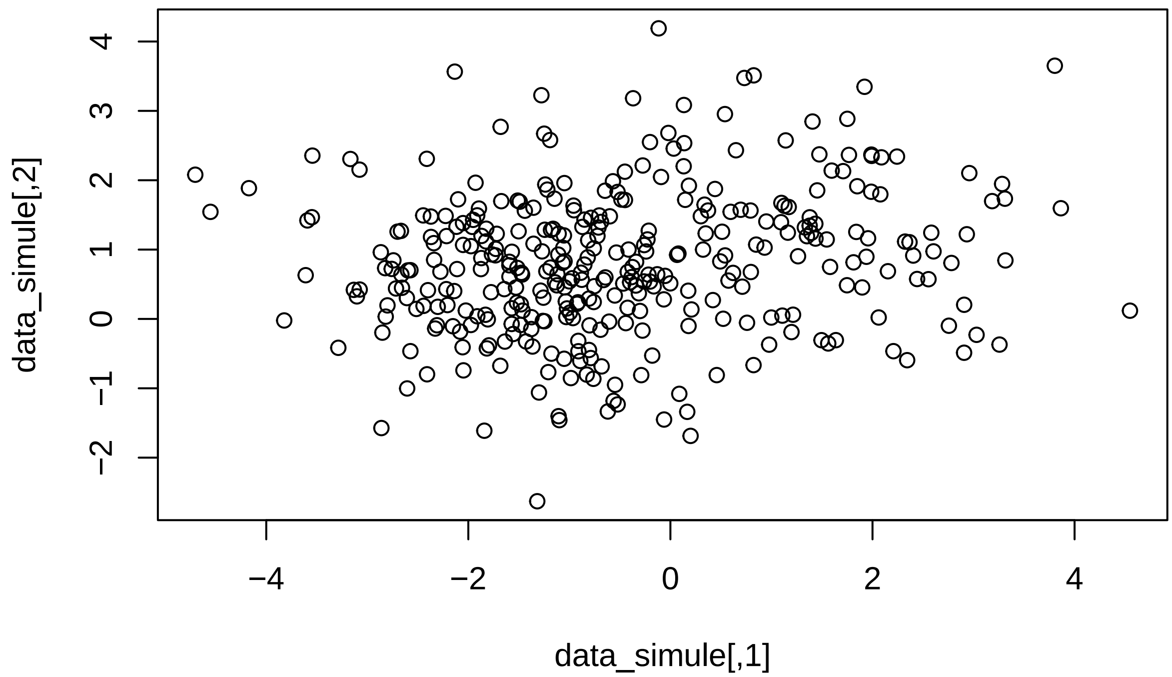
Échantillonnage d'abord N fois à partir d'une distribution multinomiale pour déterminer quel cluster génère les données, puis pour chaque résultat des échantillons de la distribution multinomiale, on utilise la distribution multi gaussienne (fonction mvnorm) pour échantillonner:

```
set.seed(1010)

K<-2 #nombre de clusters
degree <- 2  #dimensions (2D)
N <- 350 #nombre de points

pi.true <- runif(K)
pi.true <- pi.true/sum(pi.true)

mu.true <- matrix(runif(K*degree,min=-2,max=2), nrow=degree, ncol=K)

sigma.true <- array(0, dim=c(degree, degree, K))

for(k in 1:K){
  sigma.true[,,k] <- diag(1,nrow=degree, ncol=degree)
}
```

```
data_simule <-  t(apply(rmultinom(N,1,pi.true),2,function(num) {
  maxindex <- which.max(num)
  rmvnorm(1, mu.true[,maxindex], sigma.true[,,maxindex])
}))
```

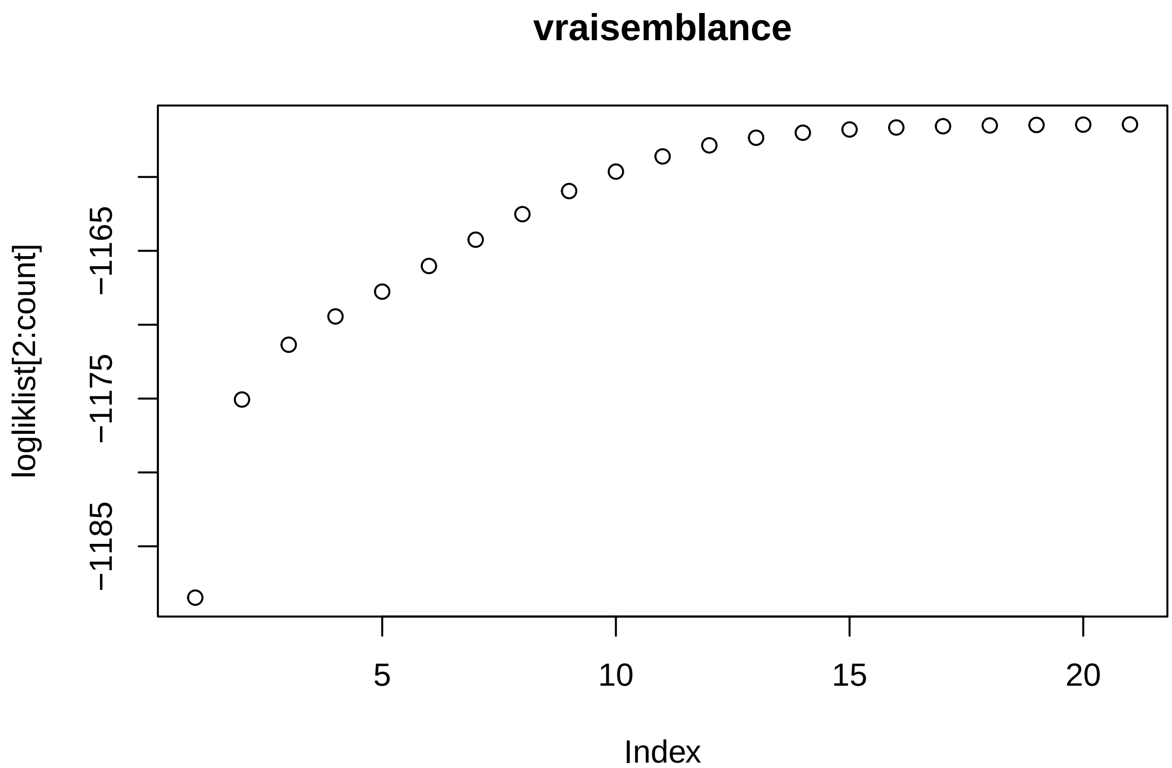
Voyons à quoi ressemblent les données:

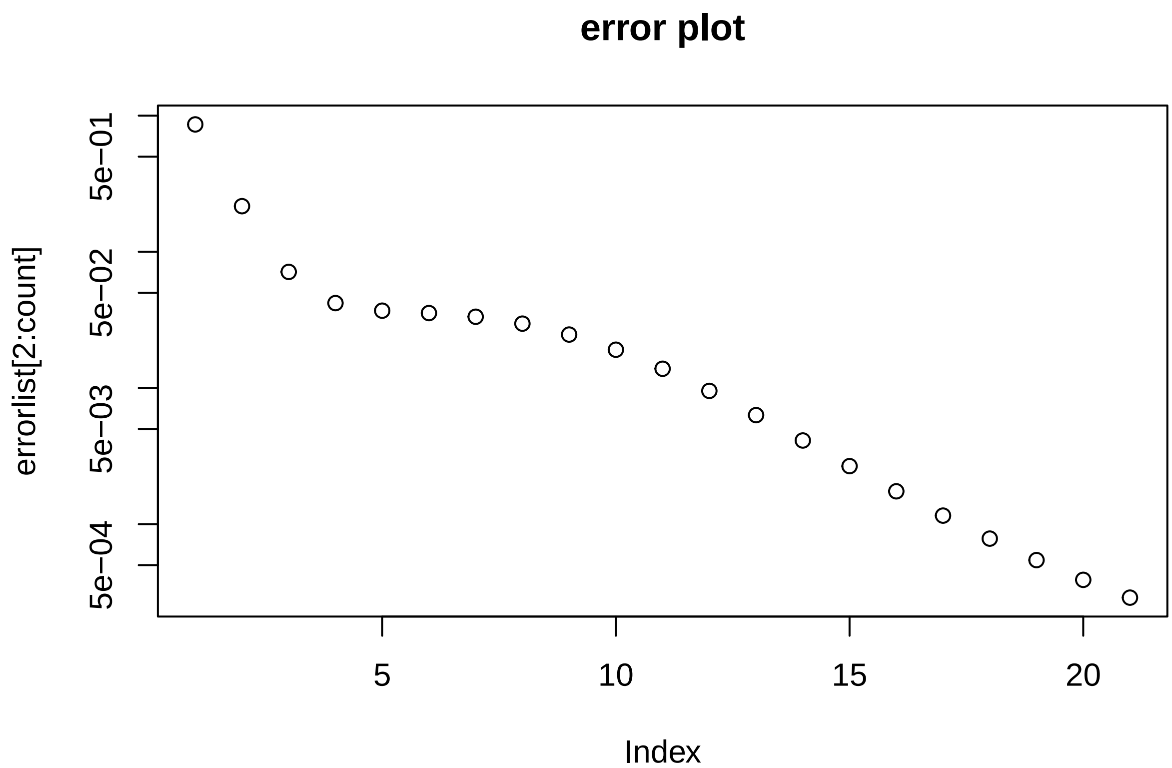
```
plot(data_simule)
```



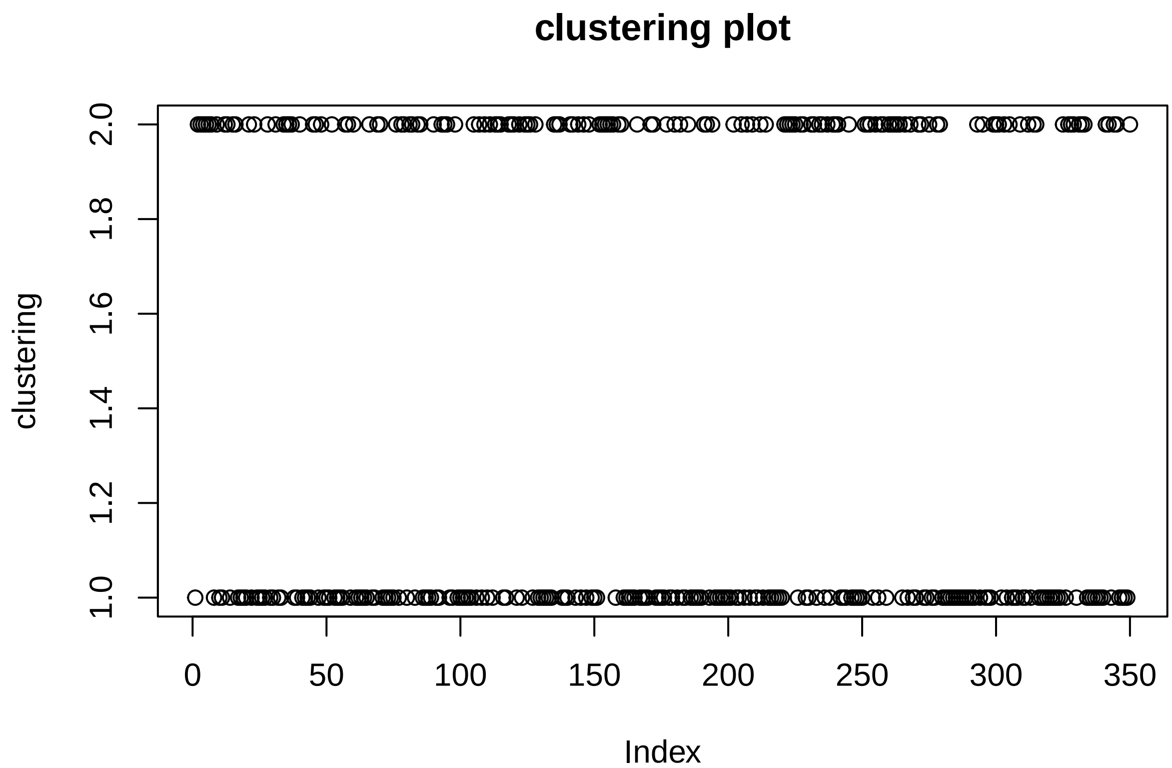

```r
n.trial<-350
e<-1e-5
res_em = EM(data_simule,K,n.trial,e)
```

```
## loglike à l'étape  1  :  -1188.474
## loglike à l'étape  2  :  -1175.066
## loglike à l'étape  3  :  -1171.355
## loglike à l'étape  4  :  -1169.441
## loglike à l'étape  5  :  -1167.761
## loglike à l'étape  6  :  -1166.026
## loglike à l'étape  7  :  -1164.245
## loglike à l'étape  8  :  -1162.517
## loglike à l'étape  9  :  -1160.952
## loglike à l'étape  10  :  -1159.637
## loglike à l'étape  11  :  -1158.609
## loglike à l'étape  12  :  -1157.86
## loglike à l'étape  13  :  -1157.344
## loglike à l'étape  14  :  -1157.004
## loglike à l'étape  15  :  -1156.787
## loglike à l'étape  16  :  -1156.651
## loglike à l'étape  17  :  -1156.565
## loglike à l'étape  18  :  -1156.512
## loglike à l'étape  19  :  -1156.479
## loglike à l'étape  20  :  -1156.458
## loglike à l'étape  21  :  -1156.444
```


vraisemblance
logliklist[2:count]
−1165
−1175
−1185
5
10
15
20
Index




```
## Cluster  1
## Pi  :
## [1] 0.5122733
## mu :
## [1] -1.45374
## Sigma :
## [1] 0.9064686
## Cluster  2
## Pi  :
## [1] 0.4877267
## mu :
## [1] 0.3949344
## Sigma :
## [1] -0.1647762
```



On remarque que l'algorithme a convrgé apres 21 étapes et que la meilleur valeur de la vraisemblance est égale à -1156.444

Les parametres éstimés pour chaque cluster sont :

#cluster1:

Pi : 0.5122733

mu : -1.45374

Sigma : 0.9064686

#cluster2:

Pi estimé : 0.4877267

mu estimé : 0.3949344

Sigma estimé : -0.1647762

```
res_em
```

```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
##                [,1]        [,2]
```

```
##   [1,] 8.452673e-01 0.15473270
##   [2,] 1.167676e-04 0.99988323
##   [3,] 6.862206e-03 0.99313779
##   [4,] 2.364459e-02 0.97635541
##   [5,] 3.622053e-04 0.99963779
##   [6,] 3.380829e-04 0.99966192
##   [7,] 1.029420e-05 0.99998971
##   [8,] 7.496405e-01 0.25035953
##   [9,] 5.207493e-02 0.94792507
##  [10,] 9.020368e-01 0.09796321
##  [11,] 8.871809e-01 0.11281914
##  [12,] 1.752861e-03 0.99824714
##  [13,] 4.558215e-01 0.54417848
##  [14,] 6.938040e-01 0.30619595
##  [15,] 4.918687e-01 0.50813129
##  [16,] 9.227317e-03 0.99077268
##  [17,] 9.119388e-01 0.08806116
##  [18,] 7.598900e-01 0.24010996
##  [19,] 7.557204e-01 0.24427958
##  [20,] 5.469484e-01 0.45305164
##  [21,] 4.552682e-02 0.95447318
##  [22,] 5.415381e-01 0.45846188
##  [23,] 1.129955e-05 0.99998870
##  [24,] 9.306382e-01 0.06936185
##  [25,] 9.299752e-01 0.07002477
##  [26,] 8.094186e-01 0.19058143
##  [27,] 7.816200e-01 0.21838004
##  [28,] 6.112977e-02 0.93887023
##  [29,] 8.577102e-01 0.14228976
##  [30,] 5.473392e-01 0.45266080
##  [31,] 4.191798e-01 0.58082015
##  [32,] 8.991233e-01 0.10087672
##  [33,] 9.105445e-01 0.08945547
##  [34,] 8.282660e-02 0.91717340
##  [35,] 1.617444e-01 0.83825559
##  [36,] 1.931631e-01 0.80683694
##  [37,] 4.585758e-03 0.99541424
##  [38,] 7.538975e-01 0.24610247
##  [39,] 7.350801e-01 0.26491986
##  [40,] 2.590771e-01 0.74092285
##  [41,] 8.245259e-01 0.17547408
##  [42,] 8.680583e-01 0.13194170
##  [43,] 5.162602e-01 0.48373984
##  [44,] 8.898643e-01 0.11013570
##  [45,] 7.290759e-03 0.99270924
##  [46,] 7.049136e-03 0.99295086
##  [47,] 8.950615e-01 0.10493852
##  [48,] 4.479932e-01 0.55200683
##  [49,] 8.883163e-01 0.11168375
##  [50,] 8.222694e-01 0.17773065
##  [51,] 8.422362e-01 0.15776377
##  [52,] 4.572945e-01 0.54270553
##  [53,] 8.877791e-01 0.11222087
##  [54,] 8.504971e-01 0.14950294
```

```
##  [55,] 8.839871e-01 0.11601286
##  [56,] 8.082511e-01 0.19174889
##  [57,] 1.103642e-07 0.99999989
##  [58,] 4.979733e-04 0.99950203
##  [59,] 8.307439e-01 0.16925605
##  [60,] 9.780957e-03 0.99021904
##  [61,] 9.158010e-01 0.08419905
##  [62,] 7.262499e-01 0.27375015
##  [63,] 7.627757e-01 0.23722429
##  [64,] 6.869960e-01 0.31300401
##  [65,] 5.386428e-01 0.46135721
##  [66,] 4.766668e-04 0.99952333
##  [67,] 8.997386e-01 0.10026136
##  [68,] 8.839299e-01 0.11607012
##  [69,] 4.761177e-01 0.52388232
##  [70,] 3.466247e-01 0.65337531
##  [71,] 8.275565e-01 0.17244346
##  [72,] 8.710766e-01 0.12892340
##  [73,] 9.089556e-01 0.09104436
##  [74,] 5.024765e-01 0.49752352
##  [75,] 8.673701e-01 0.13262994
##  [76,] 3.375448e-02 0.96624552
##  [77,] 8.118202e-01 0.18817984
##  [78,] 4.667734e-01 0.53322655
##  [79,] 2.480428e-01 0.75195723
##  [80,] 7.866395e-01 0.21336045
##  [81,] 1.824608e-01 0.81753917
##  [82,] 1.282558e-01 0.87174416
##  [83,] 8.497978e-01 0.15020224
##  [84,] 4.942225e-01 0.50577753
##  [85,] 2.118595e-03 0.99788140
##  [86,] 5.439238e-01 0.45607619
##  [87,] 6.996753e-01 0.30032474
##  [88,] 8.730552e-01 0.12694483
##  [89,] 8.563563e-01 0.14364366
##  [90,] 1.040159e-01 0.89598410
##  [91,] 9.084120e-01 0.09158804
##  [92,] 5.537923e-01 0.44620774
##  [93,] 5.376515e-02 0.94623485
##  [94,] 1.219146e-03 0.99878085
##  [95,] 2.493426e-01 0.75065741
##  [96,] 8.323612e-01 0.16763885
##  [97,] 9.227563e-01 0.07724368
##  [98,] 8.492136e-04 0.99915079
##  [99,] 9.179126e-01 0.08208741
## [100,] 7.040620e-01 0.29593803
## [101,] 8.728552e-01 0.12714476
## [102,] 9.367903e-01 0.06320971
## [103,] 8.380440e-01 0.16195600
## [104,] 9.077299e-01 0.09227010
## [105,] 1.140519e-01 0.88594807
## [106,] 6.070968e-01 0.39290325
## [107,] 1.879795e-03 0.99812020
## [108,] 5.494964e-01 0.45050363
```

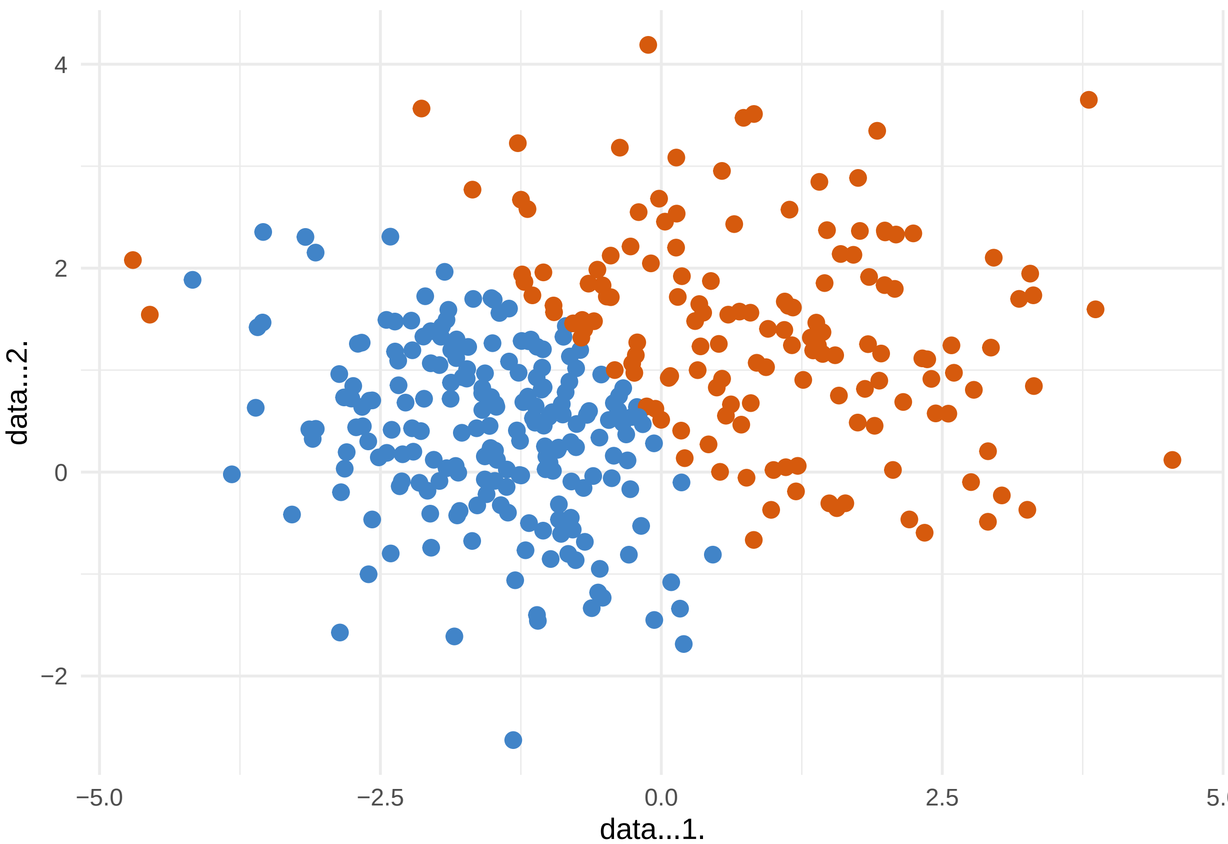
```
## [109,] 1.326604e-01 0.86733963
## [110,] 5.981458e-01 0.40185418
## [111,] 3.969138e-01 0.60308618
## [112,] 7.630340e-01 0.23696602
## [113,] 1.916642e-01 0.80833577
## [114,] 6.705513e-04 0.99932945
## [115,] 6.127970e-03 0.99387203
## [116,] 7.503045e-01 0.24969555
## [117,] 7.870779e-01 0.21292213
## [118,] 3.751801e-02 0.96248199
## [119,] 4.826245e-04 0.99951738
## [120,] 9.865376e-05 0.99990135
## [121,] 9.258021e-01 0.07419790
## [122,] 2.521932e-01 0.74780682
## [123,] 9.212855e-01 0.07871455
## [124,] 2.838328e-01 0.71616722
## [125,] 8.948762e-02 0.91051238
## [126,] 1.671302e-02 0.98328698
## [127,] 5.641267e-01 0.43587325
## [128,] 4.935132e-01 0.50648677
## [129,] 8.932792e-01 0.10672076
## [130,] 7.691603e-01 0.23083970
## [131,] 7.881234e-01 0.21187658
## [132,] 8.528569e-01 0.14714309
## [133,] 8.228386e-01 0.17716135
## [134,] 9.296095e-01 0.07039052
## [135,] 1.495175e-01 0.85048246
## [136,] 3.010478e-01 0.69895216
## [137,] 8.772924e-02 0.91227076
## [138,] 8.950043e-01 0.10499565
## [139,] 8.319105e-01 0.16808946
## [140,] 5.954374e-01 0.40456258
## [141,] 2.181378e-01 0.78186219
## [142,] 4.498028e-02 0.95501972
## [143,] 5.985406e-01 0.40145944
## [144,] 4.453874e-01 0.55461256
## [145,] 8.919260e-01 0.10807401
## [146,] 4.290108e-01 0.57098923
## [147,] 7.614897e-01 0.23851029
## [148,] 4.263466e-02 0.95736534
## [149,] 8.217290e-01 0.17827099
## [150,] 8.729214e-01 0.12707861
## [151,] 6.679135e-01 0.33208650
## [152,] 8.383150e-03 0.99161685
## [153,] 6.283902e-03 0.99371610
## [154,] 1.088072e-02 0.98911928
## [155,] 6.813771e-05 0.99993186
## [156,] 1.330368e-04 0.99986696
## [157,] 3.791704e-01 0.62082957
## [158,] 5.039092e-01 0.49609080
## [159,] 2.606973e-04 0.99973930
## [160,] 1.439471e-01 0.85605290
## [161,] 8.131300e-01 0.18686998
## [162,] 6.641593e-01 0.33584071
```

```
## [163,] 7.435479e-01 0.25645207
## [164,] 6.536802e-01 0.34631980
## [165,] 8.744965e-01 0.12550351
## [166,] 1.559998e-02 0.98440002
## [167,] 7.832645e-01 0.21673554
## [168,] 9.432357e-01 0.05676431
## [169,] 8.835984e-01 0.11640164
## [170,] 8.781497e-01 0.12185034
## [171,] 4.563841e-02 0.95436159
## [172,] 2.212613e-02 0.97787387
## [173,] 8.591686e-01 0.14083136
## [174,] 9.275826e-01 0.07241741
## [175,] 8.068903e-01 0.19310974
## [176,] 8.443728e-01 0.15562719
## [177,] 2.620116e-04 0.99973799
## [178,] 9.197289e-01 0.08027110
## [179,] 9.109783e-01 0.08902167
## [180,] 2.902209e-01 0.70977909
## [181,] 9.055280e-01 0.09447199
## [182,] 1.839582e-04 0.99981604
## [183,] 8.269447e-01 0.17305526
## [184,] 9.050539e-01 0.09494611
## [185,] 3.129285e-03 0.99687072
## [186,] 6.426905e-01 0.35730948
## [187,] 6.083958e-01 0.39160423
## [188,] 7.686469e-01 0.23135309
## [189,] 7.284082e-01 0.27159181
## [190,] 7.686337e-01 0.23136633
## [191,] 1.486959e-03 0.99851304
## [192,] 2.575580e-03 0.99742442
## [193,] 6.311307e-01 0.36886932
## [194,] 8.412925e-03 0.99158708
## [195,] 8.992545e-01 0.10074552
## [196,] 9.190813e-01 0.08091869
## [197,] 9.320704e-01 0.06792955
## [198,] 8.728849e-01 0.12711514
## [199,] 8.209691e-01 0.17903088
## [200,] 7.364219e-01 0.26357811
## [201,] 7.716832e-01 0.22831685
## [202,] 6.370127e-04 0.99936299
## [203,] 9.310273e-01 0.06897271
## [204,] 6.434968e-01 0.35650319
## [205,] 3.078364e-06 0.99999692
## [206,] 5.441128e-01 0.45588719
## [207,] 3.887150e-05 0.99996113
## [208,] 7.418044e-01 0.25819556
## [209,] 2.793310e-05 0.99997207
## [210,] 8.903729e-01 0.10962706
## [211,] 5.910579e-01 0.40894206
## [212,] 1.088337e-01 0.89116635
## [213,] 5.074071e-01 0.49259287
## [214,] 7.322526e-04 0.99926775
## [215,] 9.072191e-01 0.09278085
## [216,] 7.974329e-01 0.20256707
```

```
## [217,] 9.109503e-01 0.08904975
## [218,] 6.327055e-01 0.36729454
## [219,] 7.162864e-01 0.28371358
## [220,] 8.207788e-01 0.17922120
## [221,] 1.779783e-02 0.98220217
## [222,] 1.057114e-06 0.99999894
## [223,] 3.012997e-01 0.69870031
## [224,] 2.693798e-01 0.73062016
## [225,] 1.627281e-01 0.83727185
## [226,] 8.745260e-01 0.12547397
## [227,] 3.105432e-01 0.68945684
## [228,] 9.575573e-02 0.90424427
## [229,] 8.890343e-01 0.11096569
## [230,] 9.294929e-01 0.07050714
## [231,] 1.545276e-01 0.84547240
## [232,] 9.456476e-02 0.90543524
## [233,] 9.017513e-01 0.09824870
## [234,] 2.750210e-04 0.99972498
## [235,] 7.026568e-02 0.92973432
## [236,] 8.194219e-01 0.18057807
## [237,] 7.367397e-05 0.99992633
## [238,] 8.098776e-01 0.19012243
## [239,] 4.844341e-01 0.51556586
## [240,] 5.794204e-03 0.99420580
## [241,] 8.848273e-03 0.99115173
## [242,] 9.066227e-01 0.09337732
## [243,] 6.931918e-01 0.30680825
## [244,] 5.398683e-01 0.46013166
## [245,] 1.040165e-02 0.98959835
## [246,] 9.199997e-01 0.08000029
## [247,] 5.106819e-01 0.48931806
## [248,] 9.122627e-01 0.08773725
## [249,] 8.349542e-01 0.16504576
## [250,] 9.203077e-01 0.07969229
## [251,] 9.886378e-02 0.90113622
## [252,] 4.717691e-04 0.99952823
## [253,] 3.933922e-01 0.60660780
## [254,] 5.695681e-01 0.43043190
## [255,] 4.712762e-01 0.52872382
## [256,] 8.811828e-01 0.11881716
## [257,] 5.095757e-10 1.00000000
## [258,] 3.256470e-01 0.67435300
## [259,] 9.093094e-01 0.09069063
## [260,] 4.045859e-01 0.59541409
## [261,] 7.459629e-02 0.92540371
## [262,] 2.207269e-02 0.97792731
## [263,] 1.238114e-04 0.99987619
## [264,] 4.108708e-02 0.95891292
## [265,] 7.592809e-01 0.24071909
## [266,] 2.166692e-01 0.78333076
## [267,] 8.068520e-01 0.19314803
## [268,] 6.685874e-04 0.99933141
## [269,] 8.150279e-01 0.18497212
## [270,] 9.308557e-01 0.06914428
```

```
## [271,] 4.674126e-01 0.53258744
## [272,] 4.210378e-04 0.99957896
## [273,] 9.259629e-01 0.07403713
## [274,] 7.185693e-01 0.28143070
## [275,] 1.562223e-06 0.99999844
## [276,] 8.378825e-01 0.16211748
## [277,] 8.730734e-01 0.12692661
## [278,] 2.670115e-02 0.97329885
## [279,] 4.693884e-01 0.53061165
## [280,] 6.028301e-01 0.39716994
## [281,] 8.940782e-01 0.10592181
## [282,] 6.998572e-01 0.30014277
## [283,] 7.304841e-01 0.26951588
## [284,] 8.773259e-01 0.12267414
## [285,] 8.830583e-01 0.11694170
## [286,] 6.130460e-01 0.38695397
## [287,] 7.963577e-01 0.20364229
## [288,] 7.660119e-01 0.23398807
## [289,] 8.710637e-01 0.12893633
## [290,] 7.697544e-01 0.23024557
## [291,] 8.840160e-01 0.11598397
## [292,] 7.965364e-01 0.20346358
## [293,] 8.307209e-03 0.99169279
## [294,] 8.189247e-01 0.18107533
## [295,] 2.069087e-01 0.79309127
## [296,] 7.169559e-01 0.28304407
## [297,] 9.343806e-01 0.06561945
## [298,] 8.976677e-01 0.10233233
## [299,] 5.810384e-03 0.99418962
## [300,] 6.260606e-02 0.93739394
## [301,] 7.304477e-03 0.99269552
## [302,] 5.887785e-01 0.41122145
## [303,] 2.222949e-04 0.99977771
## [304,] 9.069553e-01 0.09304468
## [305,] 4.459181e-01 0.55408188
## [306,] 8.911286e-01 0.10887139
## [307,] 6.022328e-01 0.39776720
## [308,] 8.767467e-01 0.12325326
## [309,] 7.889057e-02 0.92110943
## [310,] 6.496354e-01 0.35036458
## [311,] 6.505233e-01 0.34947668
## [312,] 2.307534e-02 0.97692466
## [313,] 6.386312e-01 0.36136879
## [314,] 1.415002e-07 0.99999986
## [315,] 3.241719e-06 0.99999676
## [316,] 8.423579e-01 0.15764215
## [317,] 7.172942e-01 0.28270585
## [318,] 8.651227e-01 0.13487728
## [319,] 8.053887e-01 0.19461132
## [320,] 8.228091e-01 0.17719090
## [321,] 8.817867e-01 0.11821330
## [322,] 7.835878e-01 0.21641218
## [323,] 8.573623e-01 0.14263770
## [324,] 6.215655e-01 0.37843452
```

```
## [325,] 8.514677e-03 0.99148532
## [326,] 8.295673e-01 0.17043270
## [327,] 1.047573e-01 0.89524267
## [328,] 9.114954e-03 0.99088505
## [329,] 2.289794e-04 0.99977102
## [330,] 8.086375e-01 0.19136246
## [331,] 3.286688e-02 0.96713312
## [332,] 7.071091e-04 0.99929289
## [333,] 3.562987e-01 0.64370135
## [334,] 7.979868e-01 0.20201320
## [335,] 7.340717e-01 0.26592833
## [336,] 8.624337e-01 0.13756628
## [337,] 7.450273e-01 0.25497270
## [338,] 5.564919e-01 0.44350809
## [339,] 8.731116e-01 0.12688836
## [340,] 7.854930e-01 0.21450696
## [341,] 6.252023e-03 0.99374798
## [342,] 3.511369e-01 0.64886306
## [343,] 9.085369e-01 0.09146309
## [344,] 3.379459e-04 0.99966205
## [345,] 4.558573e-04 0.99954414
## [346,] 6.854344e-01 0.31456560
## [347,] 8.862438e-01 0.11375623
## [348,] 7.778825e-01 0.22211753
## [349,] 9.123170e-01 0.08768301
## [350,] 1.018101e-04 0.99989819
##
## [[4]]
##   [1] 1 2 2 2 2 2 2 1 2 1 1 2 2 1 2 2 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2 1 1 2 2 2 2
##  [38] 1 1 2 1 1 1 1 2 2 1 2 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1
##  [75] 1 2 1 2 2 1 2 2 1 2 2 1 1 1 1 2 1 1 2 2 2 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 2
## [112] 1 2 2 2 1 1 2 2 2 1 2 1 2 2 2 1 2 1 1 1 1 1 1 2 2 2 1 1 1 2 2 1 2 1 2 1 2
## [149] 1 1 1 2 2 2 2 2 2 1 2 2 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1 2 1 1 2
## [186] 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 2 1 1 2 1 2 1 1 1 1 1 1 2 2
## [223] 2 2 2 1 2 2 1 1 2 2 1 2 2 1 2 1 2 2 2 1 1 1 2 1 1 1 1 1 2 2 2 1 2 1 2 2 1
## [260] 2 2 2 2 2 1 2 1 2 1 1 2 2 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1
## [297] 1 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 2 1 2 2 2 1 2 2 2
## [334] 1 1 1 1 1 1 1 2 2 1 2 2 1 1 1 1 2
##
## [[5]]
## NULL
##
## [[6]]
```



On remarque ici que la probabilité que le premier point appartient au premier cluster est plus grande que la probabilité d'appartenace au second cluster

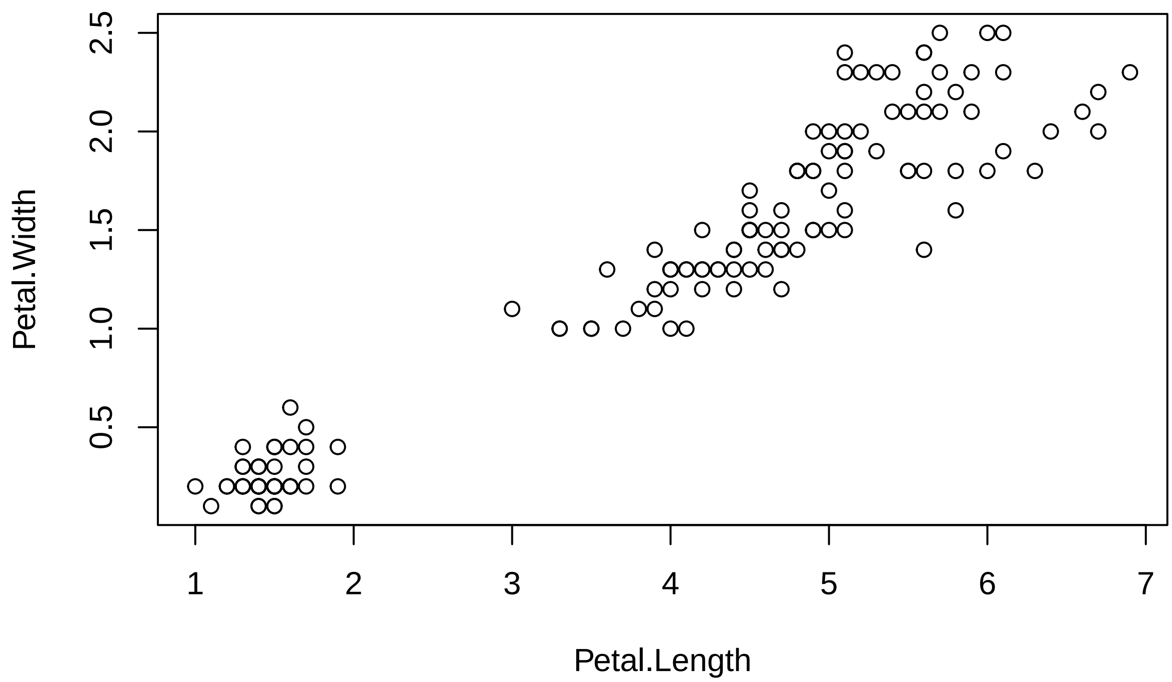
On voit bien que le premier point appartient au premier cluster

Les données sont bien regroupées en deux clusters selon le plot final

Deuxieme expérience: appliquer EM sur les données Iris

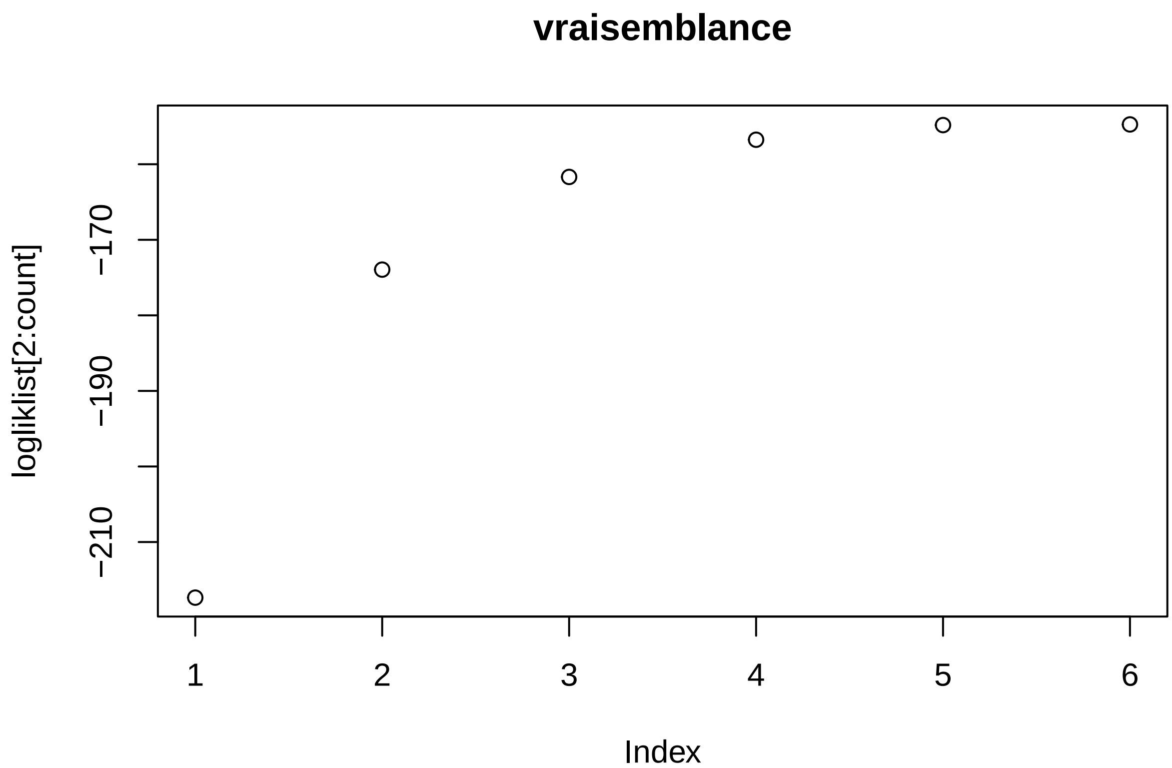
#charger les données:

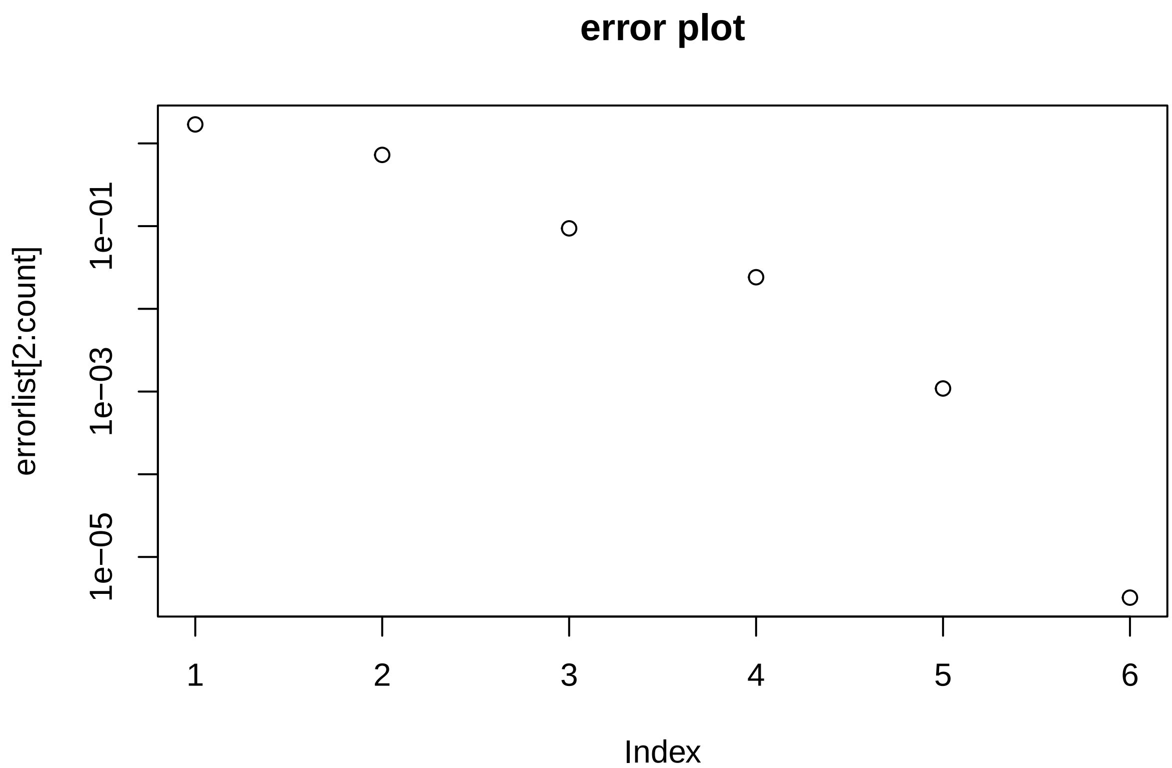
```
data_iris<-as.matrix(iris[,3:4])
plot(data_iris)
```



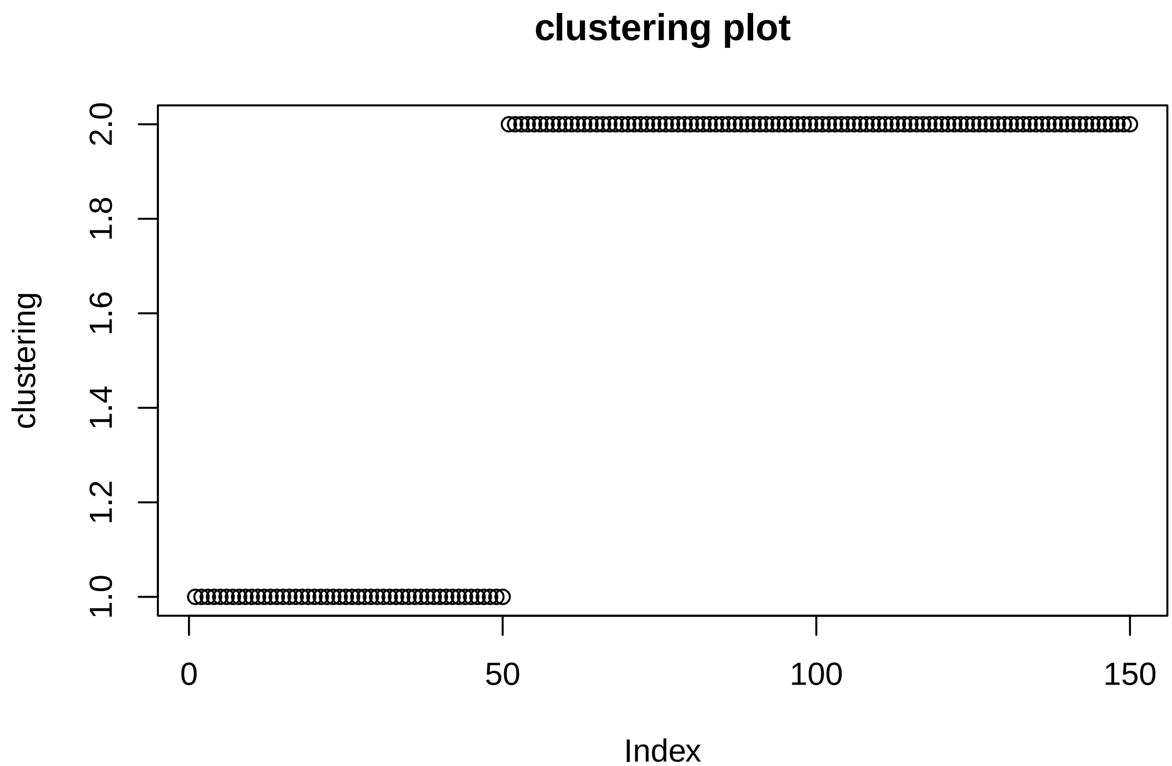

```
set.seed(1010)

K<-2  #nombre de clusters
degree <- 2  #degree de données(2D)
N <- 150 #nombre de samples
n.trial<-150
e<-1e-5
res_em = EM(data_iris,K,n.trial,e)
```

```
## loglike à l'étape  1  :  -217.3579
## loglike à l'étape  2  :  -173.9462
## loglike à l'étape  3  :  -161.6819
## loglike à l'étape  4  :  -156.7562
## loglike à l'étape  5  :  -154.8197
## loglike à l'étape  6  :  -154.7315
```


vraisemblance
logliklist[2:count]
−170
−190
−210
1
2
3
4
5
6
Index




```
## Cluster  1
## Pi  :
## [1] 0.3331373
## mu :
## [1] 1.461816
## Sigma :
## [1] 0.0295005
## Cluster  2
## Pi  :
## [1] 0.6668627
## mu :
## [1] 0.2459251
## Sigma :
## [1] 0.005936244
```



On voit que l'algorithme converge apres 6 étapes avec une meiileur vraisemblance de -154.7315

Les parametres éstimés pour chaque cluster sont :

#cluster1:

pi estimé : 0.3331373

mu estimé : 1.461816

Sigma estimé: 0.0295005

#Cluster2:

Pi estimé : 0.6668627

mu estimé : 0.2459251

Sigma estimé : 0.005936244

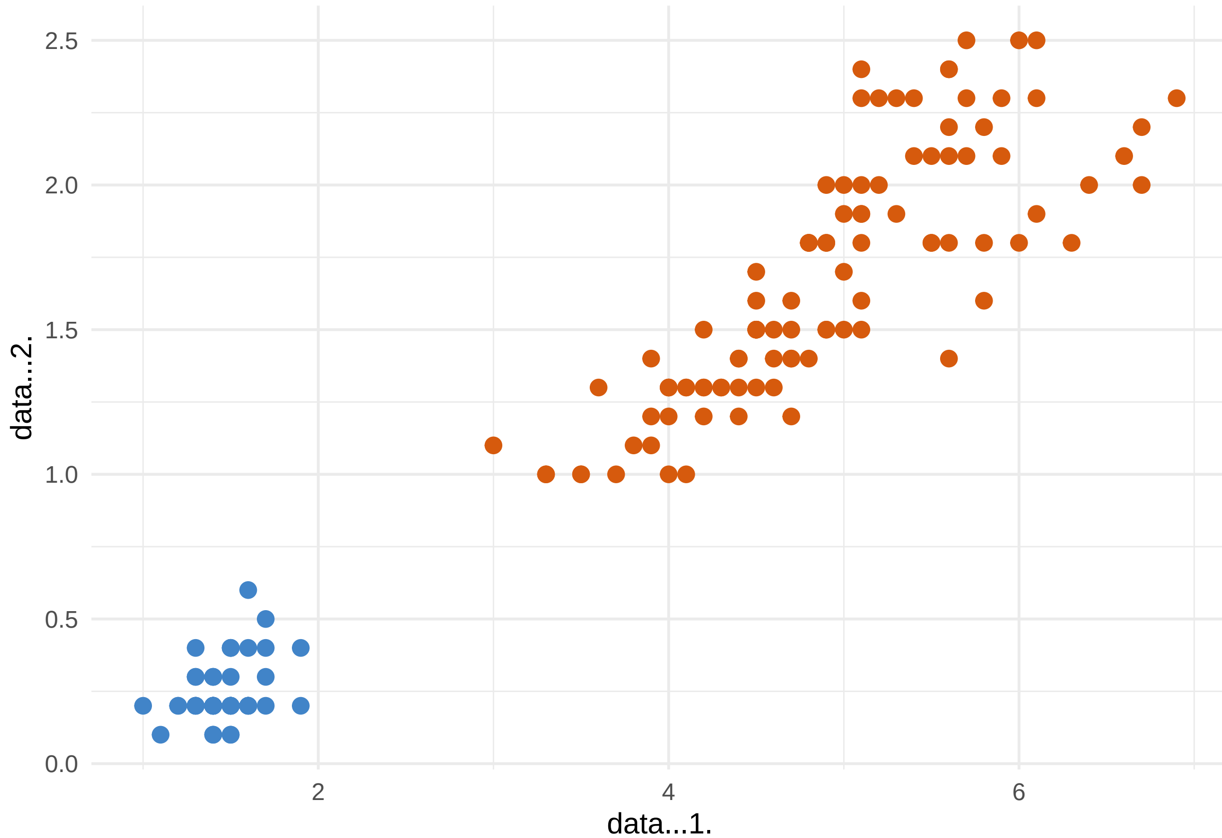
```
res_em
```

```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
##                   [,1]          [,2]
##   [1,]  9.999772e-01 2.278703e-05
```

```
##   [2,]  9.999772e-01 2.278703e-05
##   [3,]  9.999816e-01 1.837453e-05
##   [4,]  9.999605e-01 3.950937e-05
##   [5,]  9.999772e-01 2.278703e-05
##   [6,]  9.996063e-01 3.936846e-04
##   [7,]  9.999756e-01 2.442551e-05
##   [8,]  9.999605e-01 3.950937e-05
##   [9,]  9.999772e-01 2.278703e-05
##  [10,]  9.999002e-01 9.983710e-05
##  [11,]  9.999605e-01 3.950937e-05
##  [12,]  9.999042e-01 9.577267e-05
##  [13,]  9.999497e-01 5.033173e-05
##  [14,]  9.999518e-01 4.816208e-05
##  [15,]  9.999793e-01 2.071514e-05
##  [16,]  9.999179e-01 8.211331e-05
##  [17,]  9.999346e-01 6.542375e-05
##  [18,]  9.999756e-01 2.442551e-05
##  [19,]  9.997677e-01 2.323246e-04
##  [20,]  9.999630e-01 3.701740e-05
##  [21,]  9.996755e-01 3.245296e-04
##  [22,]  9.999179e-01 8.211331e-05
##  [23,]  9.999280e-01 7.195316e-05
##  [24,]  9.984221e-01 1.577901e-03
##  [25,]  9.899109e-01 1.008914e-02
##  [26,]  9.999042e-01 9.577267e-05
##  [27,]  9.998479e-01 1.520706e-04
##  [28,]  9.999605e-01 3.950937e-05
##  [29,]  9.999772e-01 2.278703e-05
##  [30,]  9.999042e-01 9.577267e-05
##  [31,]  9.999042e-01 9.577267e-05
##  [32,]  9.999179e-01 8.211331e-05
##  [33,]  9.999002e-01 9.983710e-05
##  [34,]  9.999772e-01 2.278703e-05
##  [35,]  9.999605e-01 3.950937e-05
##  [36,]  9.999793e-01 2.071514e-05
##  [37,]  9.999816e-01 1.837453e-05
##  [38,]  9.999497e-01 5.033173e-05
##  [39,]  9.999816e-01 1.837453e-05
##  [40,]  9.999605e-01 3.950937e-05
##  [41,]  9.999775e-01 2.253320e-05
##  [42,]  9.999775e-01 2.253320e-05
##  [43,]  9.999816e-01 1.837453e-05
##  [44,]  9.924674e-01 7.532621e-03
##  [45,]  9.928344e-01 7.165581e-03
##  [46,]  9.999756e-01 2.442551e-05
##  [47,]  9.999042e-01 9.577267e-05
##  [48,]  9.999772e-01 2.278703e-05
##  [49,]  9.999605e-01 3.950937e-05
##  [50,]  9.999772e-01 2.278703e-05
##  [51,]  9.683745e-83 1.000000e+00
##  [52,]  3.316115e-77 1.000000e+00
##  [53,]  4.819959e-94 1.000000e+00
##  [54,]  8.376372e-54 1.000000e+00
##  [55,]  3.384951e-81 1.000000e+00
```

```
##   [56,]  3.684204e-72 1.000000e+00
##   [57,]  2.663740e-88 1.000000e+00
##   [58,]  2.371008e-27 1.000000e+00
##   [59,]  2.873178e-76 1.000000e+00
##   [60,]  2.960414e-53 1.000000e+00
##   [61,]  1.431601e-32 1.000000e+00
##   [62,]  4.174304e-66 1.000000e+00
##   [63,]  3.639035e-48 1.000000e+00
##   [64,]  9.683745e-83 1.000000e+00
##   [65,]  9.740784e-42 1.000000e+00
##   [66,]  1.363424e-70 1.000000e+00
##   [67,]  3.316115e-77 1.000000e+00
##   [68,]  1.012466e-51 1.000000e+00
##   [69,]  3.316115e-77 1.000000e+00
##   [70,]  2.556268e-46 1.000000e+00
##   [71,]  1.823206e-99 1.000000e+00
##   [72,]  8.376372e-54 1.000000e+00
##   [73,]  4.819959e-94 1.000000e+00
##   [74,]  1.120223e-78 1.000000e+00
##   [75,]  2.216497e-64 1.000000e+00
##   [76,]  1.363424e-70 1.000000e+00
##   [77,]  4.419995e-87 1.000000e+00
##   [78,] 1.416759e-104 1.000000e+00
##   [79,]  3.316115e-77 1.000000e+00
##   [80,]  1.431601e-32 1.000000e+00
##   [81,]  4.108354e-43 1.000000e+00
##   [82,]  2.262176e-38 1.000000e+00
##   [83,]  2.950152e-48 1.000000e+00
##   [84,] 4.543008e-106 1.000000e+00
##   [85,]  3.316115e-77 1.000000e+00
##   [86,]  2.730792e-80 1.000000e+00
##   [87,]  2.471319e-85 1.000000e+00
##   [88,]  3.378925e-68 1.000000e+00
##   [89,]  3.489843e-57 1.000000e+00
##   [90,]  8.376372e-54 1.000000e+00
##   [91,]  3.536743e-66 1.000000e+00
##   [92,]  1.517465e-78 1.000000e+00
##   [93,]  1.502065e-51 1.000000e+00
##   [94,]  2.371008e-27 1.000000e+00
##   [95,]  1.039943e-60 1.000000e+00
##   [96,]  1.424750e-58 1.000000e+00
##   [97,]  1.039943e-60 1.000000e+00
##   [98,]  2.216497e-64 1.000000e+00
##   [99,]  1.053164e-22 1.000000e+00
##  [100,]  3.489843e-57 1.000000e+00
##  [101,] 2.831259e-191 1.000000e+00
##  [102,] 1.623751e-116 1.000000e+00
##  [103,] 1.548529e-165 1.000000e+00
##  [104,] 1.467780e-137 1.000000e+00
##  [105,] 1.419006e-164 1.000000e+00
##  [106,] 2.439833e-208 1.000000e+00
##  [107,]  9.497844e-84 1.000000e+00
##  [108,] 1.559846e-178 1.000000e+00
##  [109,] 1.540477e-148 1.000000e+00
```

```
## [110,] 7.281627e-197 1.000000e+00
## [111,] 9.541125e-121 1.000000e+00
## [112,] 6.366191e-126 1.000000e+00
## [113,] 2.795812e-144 1.000000e+00
## [114,] 2.547700e-116 1.000000e+00
## [115,] 2.054440e-141 1.000000e+00
## [116,] 1.264386e-144 1.000000e+00
## [117,] 2.740593e-132 1.000000e+00
## [118,] 4.427817e-219 1.000000e+00
## [119,] 4.275919e-237 1.000000e+00
## [120,]  1.287601e-98 1.000000e+00
## [121,] 3.001071e-164 1.000000e+00
## [122,] 4.865760e-112 1.000000e+00
## [123,] 2.293695e-211 1.000000e+00
## [124,] 1.019708e-103 1.000000e+00
## [125,] 1.286193e-154 1.000000e+00
## [126,] 4.231208e-160 1.000000e+00
## [127,]  1.823206e-99 1.000000e+00
## [128,] 1.019708e-103 1.000000e+00
## [129,] 2.242230e-149 1.000000e+00
## [130,] 9.110762e-143 1.000000e+00
## [131,] 2.267105e-169 1.000000e+00
## [132,] 1.547966e-191 1.000000e+00
## [133,] 4.606537e-154 1.000000e+00
## [134,] 2.460214e-103 1.000000e+00
## [135,] 4.794961e-127 1.000000e+00
## [136,] 3.341438e-186 1.000000e+00
## [137,] 1.464858e-164 1.000000e+00
## [138,] 2.740593e-132 1.000000e+00
## [139,]  1.823206e-99 1.000000e+00
## [140,] 2.493386e-139 1.000000e+00
## [141,] 1.464858e-164 1.000000e+00
## [142,] 1.098763e-135 1.000000e+00
## [143,] 1.623751e-116 1.000000e+00
## [144,] 6.190094e-175 1.000000e+00
## [145,] 2.228208e-175 1.000000e+00
## [146,] 4.407221e-140 1.000000e+00
## [147,] 4.960439e-112 1.000000e+00
## [148,] 2.555670e-125 1.000000e+00
## [149,] 2.594466e-149 1.000000e+00
## [150,] 1.167124e-112 1.000000e+00
##
## [[4]]
##   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [112] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [149] 2 2
##
## [[5]]
## NULL
##
## [[6]]
```



On remarque ici que la probabilité que le premier point appartient au premier cluster est 9.999772e-01, et pour le second cluster est 2.278703e-05

On voit bien que le premier point appartient au premier cluster!

Les données sont bien regroupées en deux clusters finalement.