Assignment_3_0523_panz2

PanZhang

3/24/2019

Derivation – E step:

$$p(z_n = k|x_n, \theta^{old}) = r_{nk}$$

$$= \frac{p(z_n = k)\, p(x_n|z_n = k)}{\sum_j p(z_n = j)\; p(x_n|z_n = j)}$$

$$= \frac{\pi_k N(x_n|\mu_k\,,\ \Sigma_k)}{\sum_j \pi_j N(x_n|\mu_j\,,\ \Sigma_j)}$$

The main question is to obtain $\pi_k N(x_n|\mu_k\,,\ \Sigma_k)$:

$$\pi_k N(x_n|\mu_k\,,\ \Sigma_k) = [\left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{|\Sigma|^{\frac{1}{2}}}\, e^{\left[-\frac{1}{2}(x_n-\mu_k)^t\Sigma^{-1}(x_n-\mu_k)\right]}]\; p_{\{z_n = k|x_n, \theta^{old}\}}$$

So, the probability linking the n'th data item to the k'th component, using

$$p(z_n = k|x_n, \theta^{old}) = \frac{p_{\{z_n = k|x_n, \theta^{old}\}} e^{\left[-\frac{1}{2}(x_n-\mu_k)^t\Sigma^{-1}(x_n-\mu_k)\right]}}{\sum_j p_{\{z_n = j|x_n, \theta^{old}\}} e^{\left[-\frac{1}{2}(x_n-\mu_j)^t\Sigma^{-1}(x_n-\mu_j)\right]}}$$

Derivation – M step:

The M step involves maximizing the expected complete data log likelihood:

$$Q(\theta, \theta^{old}) = E \textstyle\sum_n p_{\log p}(x_n, z_n|\theta)$$

$$= E \textstyle\sum_n \sum_k I(z_n = k)\log[\pi_k N(x_n|\mu_k, \Sigma_k)]$$

$$= \textstyle\sum_n p(z_n|x_n, \theta^{old})\ \ \sum_k I(z_n = k)\log[\pi_k N(x_n|\mu_k, \Sigma_k)]$$

$$= \textstyle\sum_n \sum_k r_{nk} \log[\pi_k N\,(x_n|\mu_k, \Sigma_k)]$$

$$= \sum_n \sum_k r_{nk} \log \pi_k + \sum_n \sum_k r_{nk} \log N(x_n | \mu_k, \Sigma_k)]$$

This can be optimized wrt π and μ_k, Σ_k separately. Hence we solve

$$\frac{\partial}{\partial \pi_i} [\sum_n \sum_k r_{nk} \log \pi_k + \lambda(1 - \sum_k \pi_k)] = 0$$

to find

$$\pi_k = \frac{1}{N} \sum_n r_{nk}$$

Take derivative with respect to μ_j

$$\mu_k^{new} = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$$

Take derivative with respect to Σ

$$\Sigma^{new} = \frac{\sum_n \sum_k r_{nk} (x_i - \mu_k^{new})(x_i - \mu_k^{new})^T}{\sum_n \sum_k r_{nk}}$$

Code:

```
Estep = function(data, G, para){
  pr = para$prob
  mu = para$mean
  Sinv = solve(para$Sigma)
  n = nrow(data)
  tmp = NULL
  for(k in 1:G){
    tmp = cbind(tmp,
                apply(data, 1,
                      function(x) t(x - mu[, k]) %*% Sinv %*% (x - mu[, k]))
)
  }
  tmp = -tmp/2 + matrix(log(pr), nrow=n, ncol=G, byrow=TRUE)
# tmp = tmp - apply(tmp, 1, mean)
# bigM = 15
# tmp[tmp > bigM] = bigM
# tmp[tmp < -bigM] = -bigM
  tmp = exp(tmp)
  tmp = tmp / apply(tmp, 1, sum)
```

```
  return(tmp)
}

Mstep <- function (data, G, para, post.prob ) {
  # Your Code
  # Return the updated parameters
  n = nrow(data)
  m = ncol(data)
  update.prob = apply(post.prob, 2, sum)/n
  update.mu = NULL
  update.sigma = array(0,dim=c(2,2))
  num.sigma = array(0,dim=c(2,2))
  denom.sigma = 0
  for (k in 1:G){
    num.mu = rep(0,m)
    denom = 0
    for (i in 1:n){
      num.mu = num.mu + post.prob[i,k]*data[i,]
      denom = denom + post.prob[i,k]
    }
    new.mu = num.mu/denom
    #print(new.mu)
    update.mu = cbind(update.mu,t(new.mu))
    for (i in 1:n){
      num.sigma = num.sigma + post.prob[i,k]* as.numeric(data[i,] - new.mu) %
*% t(as.numeric(data[i,] - new.mu))
    }
    denom.sigma = denom.sigma + denom
  }
  update.sigma = num.sigma/denom.sigma
  return(list(prob=update.prob, mean=update.mu, Sigma= update.sigma))
}

myEM <- function (data, T, G, para ) {
  for(t in 1: T ) {
    post.prob <- Estep(data, G, para )
    para <- Mstep (data, G, para, post.prob )
  }
  return (para)
}

library(mclust)

## Warning: package 'mclust' was built under R version 3.3.2

## Package 'mclust' version 5.4
## Type 'citation("mclust")' for citing this R package in publications.

n <- nrow(faithful)
Z <- matrix (0, n, 2)
Z[sample (1:n, 120), 1] <- 1
```

```
Z[, 2] <- 1 - Z[, 1]
ini0 <- mstep(modelName ="EEE", faithful, Z)$parameters
# Output from my EM alg
para0 <- list(prob = ini0$pro, mean=ini0$mean, Sigma = ini0$variance$Sigma)
myEM (data = faithful, G = 2, T = 10, para = para0 )
```

```
## $prob
## [1] 0.4402434 0.5597566
##
## $mean
##                   1         1
## eruptions  3.285232  3.647088
## waiting   69.266363 72.179586
##
## $Sigma
##           [,1]      [,2]
## [1,]  1.265672  13.66664
## [2,] 13.666642 182.05240
```

```
# Output from mclust
Rout <- em (modelName = "EEE", data = faithful, control = emControl(eps =0, t
ol =0, itmax = 10), parameters = ini0 )$parameters
list ( Rout$pro, Rout$mean, Rout$variance$Sigma )
```

```
## [[1]]
## [1] 0.4402434 0.5597566
##
## [[2]]
##                [,1]      [,2]
## eruptions  3.285232  3.647088
## waiting   69.266363 72.179586
##
## [[3]]
##           eruptions   waiting
## eruptions  1.265672  13.66664
## waiting   13.666642 182.05240
```