

## 情報意味論 (課題2)

慶應義塾大学理工学部  
櫻井 彰人

## レポート課題2

- レポート提出は、前回同様、web を通じて行ってください。
- レポートは電子的に作成してください。TeX, MsWord で作成して結構です。提出は pdf 形式でも結構です。
- 書くべき内容に関しては特には述べません。すべて常識的に判断してください。
- 締め切りは、4週間+ $\alpha$ 後の水曜日 i.e. 1/15 一杯とします。

(2017.12.10)

## 2-1 EM

Dempster, Laird and Rubin の論文 "Maximum Likelihood from Incomplete Data via the EM Algorithm" の p.2 に書かれている例題を、EMアルゴリズムを用いて、解いて下さい (<http://web.mit.edu/6.435/www/Dempster77.pdf>)。

プログラムを作成し、実際に動作させ、論文と同様の結果が得られるか、確認してください。なお、本講義のスライドに少し簡単にした問題の答えが書いてあります。それを参考にして結構です。

## 2.2 EM その2

$\mathbf{X}$  を、各成分が  $x_{ij}$  ( $i=1, \dots, D, j=1, \dots, M$ ) である  $D \times M$  次元のベクトル値の確率変数とする。ただし、 $x_{ij}$  の各々は  $\{0,1\}$  (0と1からなる集合) に値を取り、すべての  $i$  に対し制約条件  $\sum_j x_{ij} = 1$  を満たすものとする。 $\mathbf{X}$  の分布は、以下のように、多項分布の混合であるとする。

ただし、

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k p(\mathbf{x} | \boldsymbol{\mu}_k)$$
$$p(\mathbf{x} | \boldsymbol{\mu}_k) = \prod_{i=1}^D \prod_{j=1}^M \mu_{kij}^{x_{ij}}$$

である。パラメータ  $\mu_{kij} = p(x_{ij} = 1 | \boldsymbol{\mu}_k)$  は、 $0 \leq \mu_{kij} \leq 1$  かつ各  $k, i$  に対し制約条件  $\sum_j \mu_{kij} = 1$  を満たす。観測データ集合  $\{\mathbf{x}_n\}_{n=1, \dots, N}$  が与えられた下で、この分布の混合係数  $\pi_k$  とパラメータ  $\mu_{kij}$  を最尤推定するEMアルゴリズムのEステップとMステップ(更新式)を求めよ  
インチキな方法で結構です

## 2-3 モデル選択

Wikipedia にあるプログラム  
([http://en.wikipedia.org/wiki/File%3aEm\\_old\\_faithful.gif](http://en.wikipedia.org/wiki/File%3aEm_old_faithful.gif)) を少し手直して混合正規分布の可視化を試みる。

- プログラムを修正して、3個の正規分布、4個の正規分布の混合分布でfitしてください。よりよいfitになりますか？
  - プログラムを示してください。結果の図を回答して下さい。
- 要素分布の最適な(汎化能力が高いという意味で)個数はどのように決めたらよいでしょうか？(実際に計算しなくて結構です。考え方を述べれば結構です。)

[http://en.wikipedia.org/wiki/File%3aEm\\_old\\_faithful.gif](http://en.wikipedia.org/wiki/File%3aEm_old_faithful.gif) のうち、「#plot initial contours」の前までは、同様。これ以降を次のものに変える。  
HTMLファイル及び画像ファイルは、作業ディレクトリの下に作られる

```
#load library animation
if(!require(animation)){
  install.packages("animation")
}
library(animation)
ani.options(loop=FALSE, interval=0.25)    ##seconds per frame

saveHTML({
  ##plot initial contours
  iter <- 1
  plot.em(theta)

  #run EM and plot
  for (iter in 2:30){
    T <- E.step(theta)
    theta <- M.step(T)
    plot.em(theta)
  }
})
```