EM-Алгоритм

1 Обозначения

Пусть даны $\vec{x_1}, ..., \vec{x_n}$- точки на плоскости и k - количество кластеров
Σ_i - матрица ковариации для i-го распределения
μ_i - среднее для i-го распределения
w_{ij} - вес i-й точки для j-го распределения
$P(\vec{x_i}|j) = \frac{1}{\sqrt{2 \cdot \pi \cdot |\Sigma_j|}} \cdot \exp(-\frac{1}{2}(\vec{x_i} - \vec{\mu_j})^T {\Sigma_j}^{-1} (\vec{x_i} - \vec{\mu_j}))$ - вероятность того, что x_i лежит в i-ом распределении
$P(j|\vec{x_i}) = \frac{P(\vec{x_i}|j)}{\sum_{j=1}^{k} P(\vec{x_i}|j)}$ - вероятность того, что точка пришла из j-го распределения
$w_{ij} = \frac{P(j|\vec{x_i})}{\sum_{l=1}^{k} P(j|\vec{x_l})}$ - перерасчет весов
$\mu_{ji} = \sum_{l=1}^{n} w_{jl} \cdot x_{li}$

2 Алгоритм

Algorithm 1 EM

$\vec{w} \leftarrow random$
$\vec{\mu} \leftarrow random$
$w \leftarrow random$
while clusters changing **do**
  Е-шаг:
  for i from 1 to n **do**
  $sump_i \leftarrow 0$
  for j from 1 to k **do**
    $P(\vec{x_i}|j) \leftarrow$
    $\frac{1}{\sqrt{2 \cdot \pi \cdot |\Sigma_j|}} \cdot \exp(-\frac{1}{2}(\vec{x_i} - \vec{\mu_j})^T {\Sigma_j}^{-1} (\vec{x_i} - \vec{\mu_j}))$
    $sump_i \leftarrow sump_i + P(\vec{x_i}|j)$
  end for
  $P(j|\vec{x_i}) \leftarrow P(\vec{x_i}|j)/sump_i$
  for j from 1 to k **do**
    $w_{ij} \leftarrow \frac{P(j|\vec{x_i})}{\sum_{l=1}^{n} P(j|\vec{x_i})}$
    for l from 1 to d **do**
      $\mu_{jl} \leftarrow \mu_{jl} + w_{ji} \cdot x_{il}$
    end for
  end for
 end for
 М-шаг:
 for i from 1 to n **do**
  for j from 1 to k **do**
   for l from 1 to d **do**
    for m from 1 to d **do**
     $\Sigma_{jlm} \leftarrow \Sigma_{jlm} + w_{ji} \cdot (x_{il} - \mu_{jl}) \cdot (x_{im} - \mu_{jm})$
    end for
   end for
  end for
 end for
end while