

Probabilistic Models: Spring 2014

Poisson Mixture Model Example

We are given the following dataset \mathcal{D} of DNA sequences from a metagenomics sample. Use a Poisson mixture model (PMM) to cluster the sequences into two groups.

Index	Sequence	Count
D_1	AACCTGCCG	1
D_2	CGCGCTCAA	12
D_3	AGTGTGAGC	3
D_4	TGGGTACAC	11
D_5	GGCCGCGTG	15
D_6	CCTTAAGAG	2
D_7	GCGGAACTG	9
D_8	GCGTTGTAG	17
D_9	GTTGTAGCG	20
D_{10}	ACACGTGAC	16

Use the following initial parameters for the PMM.

- p_1 : 0.4
- p_2 : 0.6
- λ_1 : 7
- λ_2 : 8

1. Calculate the $P(k|l)$ values for D_1 , D_2 and D_3 . The others are as follows.

Index	$P(k = 1 D_i)$	$P(k = 2 D_i)$
D_4	0.2944	0.7056
D_5	0.1965	0.8035
D_6	0.5811	0.4189
D_7	0.3527	0.6473
D_8	0.1577	0.8423
D_9	0.1114	0.8886
D_{10}	0.1762	0.8238

2. Use the $P(k|l)$ values to calculate $\sum_i^N P(k|l)$ and $\sum_i^N \{P(k|l) \cdot D_i\}$ for each k .

3. Calculate the updated values for λ_k and p_k .
4. Use the new parameters to calculate $P(k|l)$ for D_1 .
5. After 5 iterations, we have the following values.
 - p_1 : 0.2997
 - p_2 : 0.7003
 - λ_1 : 2.0035
 - λ_2 : 14.2799
 - $P(k = 1|3)$: 0.9961
 - $P(k = 2|3)$: 0.0039
 - $\sum_l^N P(k = 1|l)$: 2.9974
 - $\sum_l^N P(k = 2|l)$: 7.0026
 - $\sum_l^N \{P(k = 1|l) \cdot D_l\}$: 6.0049
 - $\sum_l^N \{P(k = 2|l) \cdot D_l\}$: 99.9951

Use these values to calculate new values for λ_k and p_k . Then, use those to calculate $P(k|3)$ for each k .

Useful Equations and Algorithms

$$g(D_l : \lambda) := \frac{\lambda^{D_l}}{D_l!} \exp\{-\lambda\}$$

$$Z(k) := \sum_l^N P(k|l)$$

```

procedure PMMEM(data  $\mathcal{D}$ , initial  $p_1 \dots p_K, \lambda_1 \dots \lambda_K$ , convergence criteria  $\mathcal{C}$ )
  while  $\mathcal{C}$  has not been met do
     $q(k, l) \leftarrow p_k \cdot g(D_l, \lambda_k)$ 
     $P(k|l) \leftarrow \frac{q(k, l)}{\sum_m^K q(m, l)}$ 
     $\lambda_k \leftarrow \frac{\sum_l^N P(k|l) D_l}{Z(k)}$ 
     $p_k \leftarrow \frac{Z(k)}{N}$ 
  end while
end procedure

```

▷ Update the expectations

▷ Maximize the parameters