COMS 4771 Spring 2015

MLE mixing weights

Daniel Hsu *Warning: notes are incomplete and unchecked for correctness*

1 Maximum likelihood estimation of mixing weights

Occasionally, in the course of deriving an MLE (e.g., in the context of E-M), you may find yourself trying to find a formula for the maximizer of

$$\sum_{i=1}^{n} \sum_{j=1}^{k} w_j^{(i)} \ln \pi_j \tag{$\star$}$$

as a function of $\boldsymbol{\pi}$. Here, for each $i \in [n]$, $\boldsymbol{w}^{(i)}$ is a probability vector over $[k]$, and so is $\boldsymbol{\pi}$. This is a concave function of $\boldsymbol{\pi}$, but we have to maximize over the probability simplex.

A brain-dead way to deal with this is to just replace π_1 with $1 - \sum_{j=2}^{k} \pi_j$. So we try to maximize

$$\sum_{i=1}^{n} \left\{ w_1^{(i)} \ln \left(1 - \sum_{j=2}^{k} \pi_j \right) + \sum_{j=2}^{k} w_j^{(i)} \ln \pi_j \right\}$$

as a function of $(\pi_2, \dots, \pi_k)$. (We won't worry with any constraints for now.) For each $j \in [k] \setminus \{1\}$, the partial derivative of the above expression with respect to π_j is

$$\frac{\sum_{i=1}^{n} w_j^{(i)}}{\pi_j} - \frac{\sum_{i=1}^{n} w_1^{(i)}}{1 - \sum_{j'=2}^{k} \pi_{j'}}.$$

This is zero when

$$\left(1 - \sum_{j'=2}^{k} \pi_{j'} \right) \left(\sum_{i=1}^{n} w_j^{(i)} \right) = \pi_j \sum_{i=1}^{n} w_1^{(i)}.$$

Let us replace $1 - \sum_{j=2}^{k} \pi_j$ with π_1, and now sum both sides over all $j \in [k] \setminus \{1\}$. Using the fact that $\sum_{j=2}^{k} w_j^{(i)} = 1 - w_1^{(i)}$ for each $i \in [n]$, we have

$$\pi_1 \sum_{i=1}^{n} (1 - w_1^{(i)}) = (1 - \pi_1) \sum_{i=1}^{n} w_1^{(i)},$$

which rearranges to

$$\pi_1 = \frac{1}{n} \sum_{i=1}^{n} w_1^{(i)}.$$

Note that there was nothing special about $1 \in [k]$; we could have done all of the above where we replace the special role of $1 \in [k]$ with any $j \in [k]$. We conclude that a maximizer of $(\star)$ is

$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} w_j^{(i)}.$$

for each $j \in [k]$. It can be checked that this setting of π_j ensures that $\boldsymbol{\pi}$ is a probability vector.