BASIC PROBABILITY: THEORY
Master of Logic, University of Amsterdam, 2018 —
TEACHERS Alexandre Cremers and Jakub Dotlačil TA Bas Cornelissen

Homework problem set 6

Your homework must be handed in **electronically via Canvas before Wednesday October 12th, 22:00h**. This deadline is strict and late submissions are graded with a 0. At the end of the course, the lowest of your 7 weekly homework grades will be dropped. You are strongly encouraged to work together on the exercises, including the homework. However, after this discussion phase, you have to write down and submit your own individual solution. Numbers alone are never sufficient, always motivate your answers.

Problem 1: EM in a mixture of Poissons

Suppose you observe the data set

$$x_1^N = \{1, 5, 5, 1, 5, 7, 7, 2, 2, 5\}$$

You decide to model the data with a mixture model with 3 unobserved components, whose weights you will have to estimate in this exercise using Expectation Maximization. (As for notation; we have $N = 10$ datapoints $x_1, \dots, x_N$ which will be indexed by i and we have $M = 3$ mixture components c_1, c_2, c_3 indexed by j.)

Our model assumes that for every datapoint X_i, there is a unobserved categorical variable Y_i that takes as its value one of the three components c_1, c_2, c_3 with (unknown) probabilities w_1, w_2 and w_3 respectively. So $P(Y_i = c_j) = w_j$. For every component there exists a Poisson distribution where the observed x_i is then drawn from. The three Poisson distributions have (unknown) parameters $\lambda_{c_1}, \lambda_{c_2}$ and λ_{c_3}. We will collect all parameters in $\theta = (\lambda_{c_1}, \lambda_{c_2}, \lambda_{c_3}, w_1, w_2, w_3)$, so we can concisely summarise all this as

$$\begin{aligned}
X_i \mid Y_i = c_1, \Theta = \theta &\sim \text{Poisson}(\lambda_{c_1}) \\
X_i \mid Y_i = c_2, \Theta = \theta &\sim \text{Poisson}(\lambda_{c_2}) \\
X_i \mid Y_i = c_3, \Theta = \theta &\sim \text{Poisson}(\lambda_{c_3}) \\
Y_i \mid \Theta = \theta &\sim \text{Categorical}(w_1, w_2, w_3).
\end{aligned}$$

Furthermore, we assume independence between the mixture components. Thus, the model is

$$\begin{aligned}
&P(X_1^N = x_1^N,\ Y_1^N = y_1^N \mid \Theta = \theta) \\
&\quad = \prod_{i=1}^{N} P(Y_i = y_i \mid \Theta = \theta) \cdot P(X_i = x_i \mid Y_i = y_i, \Theta = \theta).
\end{aligned}$$

Again, the X_i are observed Poisson variables, with natural numbers as outcomes and the Y_i are unobserved categorical random variables that take values $y_i \in \{c_1, c_2, c_3\}$.

Expectation-Maximization starts from some initial guesses $\theta^{(0)}$ of the model parameters θ. We use the following:

$$\lambda_{c_1}^{(0)} = 0.5, \quad \lambda_{c_2}^{(0)} = 1, \quad \lambda_{c_3}^{(0)} = 1.5, \quad w_1^{(0)} = w_2^{(0)} = w_3^{(0)} = \frac{1}{3}$$

or equivalently $\theta^{(0)} = \big(0.5,\ 1,\ 1.5,\ \tfrac{1}{3},\ \tfrac{1}{3},\ \tfrac{1}{3}\big)$.

(a) **1pt** Compute the marginal log-likelihood of the data under the initial parameter settings, i.e. compute

$$\ln P(X_1^N = x_1^N \mid \Theta = \theta^{(0)}).$$

(b) **2pt** Compute the posterior log-probabilities for the mixture components per data point, i.e, compute

$$\ln P\big(Y_i = c_j \mid X_i = x_i, \Theta = \theta^{(0)}\big) \quad \text{for } j = 1, 2, 3$$

(c) **2pt** With the posterior probabilities in place, we can perform the E-step. Note that $\sum_{i=1}^{N} \mathbb{1}(Y_i = c_j)$ counts the number of Y_i in $Y_1, \dots, Y_N$ such that $Y_i = c_j$. Report the expected sufficient statistics for the categorical distributions, that is, compute the expected counts

$$\mathbb{E}\Big[\sum_{i=1}^{N} \mathbb{1}(Y_i = c_j) \ \mid X_1^N = x_1^n, \Theta = \theta^{(0)}\Big]$$

for $j = 1, 2, 3$.

(d) **2pt** Compute the expected sufficient statistics for the Poisson distributions, i.e. compute

$$\mathbb{E}\Big[\sum_{i=1}^{N}\big(x_i \cdot \mathbb{1}(Y_i = c_j)\big) \;\mid X_1^N = x_1^N, \Theta = \theta^{(0)}\Big]$$

for $j \in \{1, 2, 3\}$. Use a table for this exercise!

(e) **2pt** Perform the M-step and report the new parameters $\theta^{(1)}$. You can give the new $\lambda_j^{(1)}$'s for Poissons $P(X_i \mid Y_i)$ and the $w_j^{(1)}$'s for the categorical $P(Y_i)$ separately.

(f) **1pt** Compute the marginal log-likelihood of x_1^N under the new parameter settings $\theta^{(1)}$.