Sixth week practical exercises in Machine learning 1 – 2023 – Paper 1

1 Mixture models (October)

Consider a data distribution whose underlying generating process is a mixture of Poisson distributions, but we do not know the parameters of the mixture model. In this question, you are asked to derive the updated equations for the general Poisson mixture model. The Poisson distribution is:

$$P(x|\lambda) = \frac{1}{x!}\lambda^x \exp(-\lambda)$$

where $x = 0, 1, 2, ...$ (non-negative integers), $\lambda > 0$ is the 'rate' of the data; the expected value of x is λ. A mixture representation assumes the following:

$$P(x_n) = \sum_{k=1}^{K} \pi_k P(x_n|\lambda_k)$$

where $P(x_n|\lambda_k)$ is a Poisson distribution with rate λ_k and x_n is a single data observation. To answer the following questions assume we are given a dataset $\{x_1, x_2, \ldots, x_N\}$. Make sure that the constraint $\sum_k \pi_k = 1$ is satisfied (i.e. think of the log-likelihood or log-joint as f (an objective to maximize) and $\sum_k \pi_k - 1 = 0$ as $g = 0$ (a constraint that must hold)).

(a) Write down the likelihood (as usual) for the data set in terms of $\{x_1, x_2, \ldots, x_N\}$, $\{\pi_k\}$ and $\{\lambda_k\}$.

(b) Write down the log-likelihood (as usual) for the data set in terms of $\{x_1, x_2, \ldots, x_N\}$, $\{\pi_k\}$, $\{\lambda_k\}$.

(c) Let us consider the contribution of each of the Poisson components as their *responsibility* in the generating process, denoting them r_{nk} (i.e. the contribution of the kth- poisson distribution for the nth datapoint).

This responsibilities are actually the posterior probabilities given by the following expression $p(C_k|\mathbf{x}_n) = r_{nk}$

Find the expression for the responsibilities r_{nk}.

(d) Find the expression for λ_k that maximizes the log-likelihood.

(e) Find the expression for π_k that maximizes the log-likelihood.

(f) Now assume priors for π_k and λ_k, $p(\lambda_k|a, b) = \mathcal{G}(\lambda_k|a, b)$ (a Gamma prior) and $p(\pi_1, \ldots, \pi_k) = \mathcal{D}(\pi_1, \ldots, \pi_k|\alpha/K, \ldots, \alpha/K)$ (a Dirchlet distribution) respectively. These distributions are defined in the appendix of Bishop. Write down the log-joint distribution:

$$\log p(x_1, ..., x_N, \{\pi_k\}, \{\lambda_k\} | a, b, \alpha, K).$$

(g) Find the expression for λ_k that maximizes the log-joint.

(h) Find the expression for π_k that maximizes the log-joint.

(i) Write down an iterative algorithm using the above update equations (similar to the ones derived in class for the Mixture of Gaussians); include initialization and convergence check steps.