Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik

Probabilistic and Bayesian Modelling in Machine Learning and Artificial Intelligence

Manfred Opper, Théo Galy–Fajou

Summer Term 2018

Problem Sheet 2

Solutions

Problem 1 – EM algorithm for a Poisson mixture model

Consider a mixture model for a integer valued random variable $n \in \{0, 1, 2, \ldots\}$ given by the distribution

$$P(n|\boldsymbol{\theta}) = \sum_{j=1}^{M} P(j)\ P(n|\theta_j) = \sum_{j=1}^{M} P(j)\ e^{-\theta_j} \frac{\theta_j^n}{n!}\,,$$

where the component probabilities $P(n|\theta_j)$ are Poisson distributions. Based on a data set of i.i.d. samples $D = (n_1, n_2, \ldots, n_N)$ we want to estimate the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M, P(1), \ldots, P(M))$ of this mixture model.

(a) Derive an expression for the *Maximum Likelihood* estimate of θ_1 for $M = 1$, where all obervations come from the same Poisson distribution.

- Likelihood of the data set:

$$P(D|\theta_1) = \prod_{i=1}^{N} P(n_i|\theta_1) = \prod_{i=1}^{N} \exp(-\theta_1) \frac{\theta_1^{n_i}}{n_i!} = \exp(-N\theta_1) \prod_{i=1}^{N} \frac{\theta_1^{n_i}}{n_i!}$$

- Logarithm of the likelihood:

$$F = -\log P(D|\theta_1) = N\theta_1 - \sum_{i=1}^{N} n_i \log\theta_1 + \sum_{i=1}^{N} \log n_i!$$

- Calculation of the Maximum-Likelihood estimate:

$$\left.\frac{dF}{d\theta_1}\right|_{\theta_1=\theta^*} = 0 \quad \Longleftrightarrow \quad N - \sum_{i=1}^{N} \frac{n_i}{\theta^*} = 0 \quad \Longleftrightarrow \quad \theta^* = \frac{1}{N} \sum_{i=1}^{N} n_i$$

(b) For $M > 1$ the maximum likelihood estimates of the parameters are to be determined using an EM algorithm. Give explicit formulas for the update of θ_j and $P(j)$.

Hint: For the E-step (see the lecture), compute

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\theta}_t) = -\sum_{i=1}^{N}\sum_{j=1}^{M} P_t(j|n_i) \ln\left(P(n_i|\theta_j)\, P(j)\right),$$

where $P_t(j|n_i)$ is the responsibility of component j for generating data point n_i, computed with the current values of the parameters. For the M-step, minimise $\mathcal{L}$ with respect to θ_j and $P(j)$.

- Posterior for generating observation n_i from component j of the mixture model

$$P_t(j|n_i) = \left.\frac{P(j)e^{-\theta_j}\frac{\theta_j^{n_i}}{n_i!}}{\sum_{k=1}^{M} P(k)e^{-\theta_k}\frac{\theta_k^{n_i}}{n_i!}}\right|_{\theta=\theta_t} = \left.\frac{P(j)e^{-\theta_j}\theta_j^{n_i}}{\sum_{k=1}^{M} P(k)e^{-\theta_k}\theta_k^{n_i}}\right|_{\theta=\theta_t}$$

- E step:

$$\begin{aligned}
\langle\mathcal{L}\rangle &= -\sum_{i=1}^{N}\sum_{j=1}^{M} P_t(j|n_i) \ln\left(P(j)e^{-\theta_j}\frac{\theta_j^{n_i}}{n_i!}\right) \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{M} P_t(j|n_i)\left(-\theta_j + n_i\ln\theta_j - \ln n_i! + \ln P(j)\right)
\end{aligned}$$

- M step:

$$\begin{aligned}
\frac{\partial\langle\mathcal{L}\rangle}{\partial\theta_j} = 0 &\iff -\sum_{i=1}^{N} P_t(j|n_i)\left(-1 + \frac{n_i}{\theta_j}\right) = 0 \\
&\iff \theta_j = \frac{\sum_{i=1}^{N} n_i P_t(j|n_i)}{\sum_{i=1}^{N} P_t(j|n_i)} \\
\frac{\partial\langle\mathcal{L}\rangle}{\partial P(j)} = 0 &\iff -\sum_{i=1}^{N}\frac{P_t(j|n_i)}{P(j)} + \sum_{i=1}^{N}\frac{P_t(M|n_i)}{P(M)} = 0 \\
&\iff \frac{P(j)}{P(M)} = \frac{\sum_{i=1}^{N} P_t(j|n_i)}{\sum_{i=1}^{N} P_t(M|n_i)} \\
&\iff P(j) = \frac{1}{N}\sum_{i=1}^{N} P_t(j|n_i)
\end{aligned}$$

- Combined E and M step:

$$\begin{aligned}
P^*(j) &= \frac{1}{N}\sum_{i=1}^{N}\frac{P(j)e^{-\theta_j}\theta_j^{n_i}}{\sum_{k=1}^{M} P(k)e^{-\theta_k}\theta_k^{n_i}} \\
\theta_j^* &= \frac{1}{NP^*(j)}\sum_{i=1}^{N}\frac{n_i P(j)e^{-\theta_j}\theta_j^{n_i}}{\sum_{k=1}^{M} P(k)e^{-\theta_k}\theta_k^{n_i}}
\end{aligned}$$

Problem 2 – Bayesian estimation for the Poisson distribution

Consider again the Poisson distribution for an integer valued random variable $n \in \{0, 1, 2, \ldots\}$

$$P(n|\theta) = e^{-\theta} \frac{\theta^n}{n!} \ ,$$

(a) Write the Poisson distribution in the *exponential family* form

$$P(n|\theta) = f(n) \exp\left[\psi(\theta)\phi(n) + g(\theta)\right]$$

Writing

$$P(n|\theta) = \frac{1}{n!} e^{n \ln\theta - \theta}$$

wee see that $f(n) = \frac{1}{n!}$, $\phi(n) = n$, $\psi(\theta) = \ln\theta$ and $g(\theta) = -\theta$.

(b) Use this exponential family representation to show that the *conjugate prior* for the Poisson distribution is given by the *Gamma density*

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

where α, β are hyperparameters. Following the lecture, the conjugate prior is of the form

$$p(\theta) \propto \exp\left[\psi(\theta)a + bg(\theta)\right] = \theta^a e^{-b\theta}$$

for some constants a, b. To make the density normalisable, we need $a > -1$ and $\beta > 0$. Setting $\beta \equiv b$ and $\alpha = b + 1$ we get the Gamma density. For the normalisation, we have

$$\int_0^\infty \theta^{\alpha-1} e^{-\beta\theta} \, d\theta = \beta^{-\alpha} \int_0^\infty y^{\alpha-1} e^{-y} \, dy = \beta^{-\alpha}\Gamma(\alpha)$$

where the last integral gives $\Gamma(\alpha)$, the *Euler Gamma–function.*

(c) Assume that we observe Poisson data $D = (n_1, n_2, \ldots, n_N)$. Write down the posterior distribution $p(\theta|D)$ assuming the *Gamma* prior. What are *posterior mean* and MAP estimators for θ ?

The posterior distribution for θ is given by

$$p(\theta|D) = \frac{P(D|\theta)p(\theta|\alpha, \beta)}{P(D|\alpha, \beta)} \propto \prod_{i=1}^N \left(\theta^{n_i} e^{-\theta}\right) \theta^{\alpha-1} e^{-\beta\theta} = \theta^{\sum_{i=1}^N n_i + \alpha - 1} e^{-(N+\beta)\theta}$$

This is again of the *Gamma form* with parameters $\beta' \doteq N + \beta$ and $\alpha' \doteq \sum_{i=1}^N n_i + \alpha$.

The MAP estimator is the one that maximises the exponent $-\beta'\theta + (\alpha' - 1)\ln\theta$ in the posterior. Taking the derivative wrt θ yields

$$\theta_{MAP} = \frac{\alpha' - 1}{\beta'} = \frac{\sum_{i=1}^{N} n_i + \alpha - 1}{N + \beta}$$

The posterior mean is defined by

$$\theta_{mean} = \int_0^\infty \theta \; p(\theta|D) = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \int_0^\infty \theta^{\alpha'} e^{-\beta'\theta} \; d\theta = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \frac{\Gamma(\alpha' + 1)}{\beta'^{\alpha'+1}} = \frac{\alpha'}{\beta'}$$

In the last step we have used the relation $\Gamma(x + 1) = x\Gamma(x)$.

(d) Compute the *posterior variance* for large N and compare your result with the asymptotic frequentist error of the maximum likelihood estimator. **Hint:** For the computation of the frequentist error use the *Fisher Information* $J(\theta) \doteq E[(\frac{d \ln P(n|\theta)}{d\theta})^2]$ where the expectation is over the probability distribution $P(n|\theta)$.