Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 12

Exercise 1: EM Procedure with Discrete Distributions (15 + 20 + 15 P)

Consider a latent variable model composed of one Bernoulli and two Binomial distributions: The first distribution $p(z|\theta)$ produces the latent state $z \in \{H, T\}$ which can be interpreted as the head/tail outcome of a coin toss, and two distributions $p(x \mid z = H, \theta)$, $p(x \mid z = T, \theta)$ produce the observed data $x \in \{0, 1, \dots, m\}$ conditioned on outcome of the coin toss z. The probability distributions are defined as:

$$
\begin{aligned}
p(z = H \mid \theta) &= \lambda \\
p(z = T \mid \theta) &= 1 - \lambda \\
p(x \mid z = H, \theta) &= \binom{m}{x} a^x (1-a)^{m-x} \\
p(x \mid z = T, \theta) &= \binom{m}{x} b^x (1-b)^{m-x}
\end{aligned}
$$

The variable $\theta = (\lambda, a, b)$ contains the parameters of the model and these parameters need to be learned.

We draw from the model N times independently, and thus generate a dataset $\mathcal{D} = (x_1, \dots, x_N)$. The goal is now to estimate the parameters $\theta = (\lambda, a, b)$ that best explain the observed data. This can be done using expectation-maximization. Assuming all distributions of the latent variable model are independent, the data log-likelihood can be written as:

$$
p(\mathcal{D}|\theta) = \prod_{x \in \mathcal{D}} \sum_{z \in \{H,T\}} p(z|\theta) p(x|z, \theta)
$$

(a) *Show* that $p(\mathcal{D}|\theta)$ can be lower-bounded as:

$$
\log p(\mathcal{D}|\theta) \geq J(\theta) \quad \text{with} \quad J(\theta) = \sum_{x \in \mathcal{D}} \Big(G(x) + \sum_{z \in \{H,T\}} q(z|x) \big[\log p(z|\theta) + \log p(x|z, \theta) \big] \Big)
$$

where $q(z|x)$ can be any probability distribution conditioned on x, and where $G(x)$ is a quantity that does not depend on θ.

(b) Assuming some distribution $q(z|x)$, apply the maximization step of the EM procedure, in particular, show that the parameter θ that maximizes the objective $J(\theta)$ is given by:

$$
\lambda = \frac{1}{N} \sum_{x \in \mathcal{D}} q(z = H|x) \qquad a = \frac{\sum_{x \in \mathcal{D}} q(z = H|x) \cdot x}{\sum_{x \in \mathcal{D}} q(z = H|x) \cdot m} \qquad b = \frac{\sum_{x \in \mathcal{D}} q(z = T|x) \cdot x}{\sum_{x \in \mathcal{D}} q(z = T|x) \cdot m}
$$

(c) Apply the expectation step of the EM procedure that computes the new distribution $q(z|x) = p(z|x, \theta)$ for the current set of parameters θ, and show that it is given by:

$$
\begin{aligned}
q(z = H|x) &= \gamma \cdot a^x (1-a)^{m-x} \cdot \lambda \\
q(z = T|x) &= \gamma \cdot b^x (1-b)^{m-x} \cdot (1 - \lambda)
\end{aligned}
$$

with γ a normalization constant set to ensure $q(z = H|x) + q(z = T|x) = 1$.

Exercise 2: Programming (50 P)

Download the programming files on ISIS and follow the instructions.