Exercises for the course

Machine Learning 1

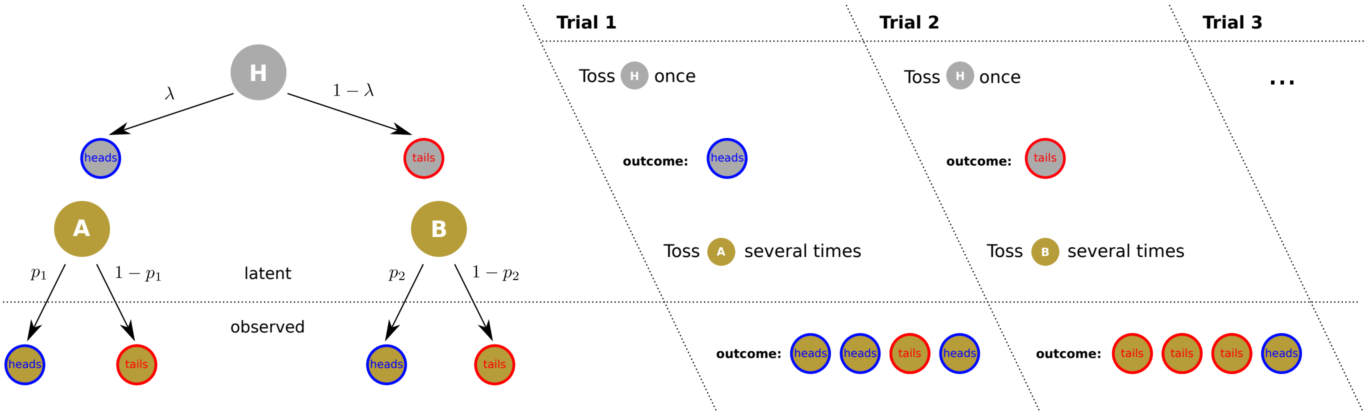
Winter semester 2018/19

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 4

Exercise 1: Discrete EM: Coin Tosses from Multiple Distributions (55 P)

Consider the following coin tossing experiment. The experimenter has three (potentially unfair) coins with different probabilities for heads and tails, one secret/hidden coin, call it coin H, and two coins that are undistinguishable in the observation, call them coin A and coin B. First, the secret coin H is tossed. If it shows heads, the experimenter tosses coin A m times, if it shows tails he tosses coin B m times. What is observed is the results from either coin A or coin B, but *not* the fact from which coin the results originated.



Mathematically, this is modelled by two random variables: Z for the secret/hidden coin H, X for the second coin which is either coin A or coin B. Z is a Bernoulli variable with

$$P(Z = \text{heads} \mid \theta) = \lambda \qquad\qquad P(Z = \text{tails} \mid \theta) = 1 - \lambda,$$

and X is an m-vector with each component an independent conditional Bernoulli variable, i.e.,

$$P(X_i = \text{heads} \mid Z = \text{heads}, \theta) = p_1 \qquad P(X_i = \text{tails} \mid Z = \text{heads}, \theta) = 1 - p_1$$
$$P(X_i = \text{heads} \mid Z = \text{tails}, \theta) = p_2 \qquad P(X_i = \text{tails} \mid Z = \text{tails}, \theta) = 1 - p_2$$

for each $1 \leq i \leq m$, where the components X_i are assumed independent, and where p_1 is the heads-probability of coin A, and p_2 the heads-probability of coin B. X is the generative observable (= variable that is observed) for the experiment.

Now, the experiment above is repeated N times. That is, coin H is tossed N times, and each time coin A or B is tossed m times. In total, $m \cdot N$ tosses are observed in the sample. We model this by a N-tuple of independent copies $\mathcal{X} := (X^{(1)}, \ldots, X^{(N)})$ of X, which form the statistical sample.

The goal is now to estimate the parameter vector $\theta = (\lambda, p_1, p_2)$ which contains all information on the experiment. This cannot be done directly, since λ is a hidden parameter, corresponding to the latent (=lat. hidden) variable Z, but it can be done by Expectation Maximization. Due to independence, the joint distribution of the observable X and latent variables Z for a fixed sample $x = (x^{(1)}, \ldots, x^{(N)})$ (an N-tuple of m-vectors with heads or tails), and for the results of the hidden coin $z = (z^{(1)}, \ldots, z^{(N)})$ (an N-tuple of heads or tails), is

$$P(\mathcal{X} = x, Z = z \mid \theta) = \prod_{i=1}^{N} P(Z = z^{(i)} \mid \theta) \prod_{j=1}^{m} P(X_j = x_j^{(i)} \mid Z = z^{(i)}, \theta),$$

The M-step of EM determines the new parameters θ^{new} given the values of the old parameters θ^{old} by solving the optimization problem

$$\theta^{\text{new}} = \text{argmax}_\theta \, Q(\theta, \theta^{\text{old}}) \tag{1}$$

with

$$Q(\theta, \theta^{\text{old}}) = \sum_{z \in \{\text{heads}, \text{tails}\}^N} P(Z = z \mid \mathcal{X} = x, \theta^{\text{old}}) \log P(\mathcal{X} = x, Z = z \mid \theta).$$

Derive the closed-form solutions for the M-step (Equation 1). That is, derive an expression for each element of the new parameter vector $\theta^{\text{new}} = (\hat{\lambda}, \hat{p}_1, \hat{p}_2)$.

Exercise 2: Programming (45 P)

Download the programming files on ISIS and follow the instructions.