

# Maschinelles Lernen 1

Wintersemester 2009/2010

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

## Blatt 15

Abgabe bis **Mittwoch, den 10. Februar 2010, 14 Uhr** im Sekretariat bei Frau Gerdes  
 (Raum FR 6052).

### Model und Notation

In this assignment, we want to take a closer look at the coin tossing example from the lecture. We have observed  $n$  sequences of coin tosses  $x^1, \dots, x^n$ , each consisting of  $m$  coin tosses, i.e.  $x^i \in \{T, H\}^m$ . The parameters to be estimated are  $\theta = (\lambda, p_1, p_2)$  corresponding to the three coins. We model the unobserved outcome of the coin toss which determines the particular coin to be used for each sequence using a latent variable  $Z$  which follows a Bernoulli distribution,

$$p(Z = H|\theta) = \lambda \quad p(Z = T|\theta) = 1 - \lambda.$$

The process which generated the sequences  $x^i$  is modeled using a random variable  $X$  conditioned on the choice of the coin,

$$\begin{aligned} p(X = H|Z = H, \theta) &= p_1 & p(X = T|Z = H, \theta) &= 1 - p_1 \\ p(X = H|Z = T, \theta) &= p_2 & p(X = T|Z = T, \theta) &= 1 - p_2. \end{aligned}$$

Since the coin tosses in the sequences  $x^i$  are independent of each other and because the tosses of the coin preceding each sequence (represented by the random variable  $Z$ ) are independent, the joint distribution of the observable  $X$  and latent variables  $Z$  is

$$p(X = x^1, \dots, x^n, Z = z|\theta) = \prod_{i=1}^n p(Z = z_i|\theta) \prod_{j=1}^m p(X = x_j^i|Z = z_i, \theta).$$

In the M-step, we determine the new parameters  $\theta^{\text{new}}$  given the values of the old parameters  $\theta^{\text{old}}$  by solving the optimization problem

$$\theta^{\text{new}} = \underset{\theta}{\operatorname{argmax}} \ Q(\theta, \theta^{\text{old}}) \quad (1)$$

with

$$Q(\theta, \theta^{\text{old}}) = \sum_{z \in \{T, H\}^n} p(Z = z|X = x^1, \dots, x^n, \theta^{\text{old}}) \log p(X = x^1, \dots, x^n, Z = z|\theta).$$

### Aufgaben

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

- Assume we have observed the following dataset

$$\begin{aligned} x^1 &= (T, T, T, T) \\ x^2 &= (H, H, H, H) \\ x^3 &= (T, T, T, H) \\ x^4 &= (H, H, H, H), \end{aligned}$$

and let  $\lambda = 0.6, p_1 = 0.4, p_2 = 0.8$  be the initial guess for the parameters. Starting from these values perform the EM-Algorithm. Write down a table showing the current estimates of the parameters  $\lambda, p_1, p_2$  in each iteration.

**(15 Punkte)**

---

Für Fragen zum Übungsblatte bitte in der Google Group <http://groups.google.com/group/mikiobraun-lehre> registrieren und die Frage an die Mailingliste stellen.