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

Machine Learning 1

Winter semester 2016/17

Group APXNLE

Exercise 7

Members

Jing Li[387272] jing.li.1@campus.tu-berlin.de

Kumar Awanish[386697] k.awanish@campus.tu-berlin.de

Manjiao Xu[386498] manjiao.xu@campus.tu-berlin.de

Rudresha Gulaganjihalli Parameshappa[386642]

Gulaganjihalliparameshappa@ campus.tu-berlin.de

Maximilian Ernst[364862] maximilian.ernst@campus.tu-berlin.de

Exercise 1: Discrete EM: Coin Tosses from Multiple Distributions

The object function is:

$$Q(\theta, \theta^{old}) = \sum_{z} P(z = z \mid X = x, \theta^{old}) \log P(X = x, z = z \mid \theta)$$

$$= \sum_{z} P(z = z \mid X = x, \theta^{old}) \sum_{i=1}^{N} \left(\log P(z = z^{(i)} \mid \theta) + \sum_{j=1}^{m} \log P(X_j = x_j^{(i)} \mid z = z^{(i)}, \theta) \right)$$

$$= \sum_{z} P(z = z \mid X = x, \theta^{old}) \sum_{z = head} \left[\log \lambda + h(x_i) \log p_1 + t(x_i) \log(1 - p_1) \right] + \sum_{z = head} \left[\log(1-\lambda) + h(x_i) \log p_2 + t(x_i) \log(1 - p_2) \right]$$

($h(x_i)$ is the number of heads when tossing coin M times $t(x_i)$ is the number of tails when tossing coin M times. So $M = h(x_i) + t(x_i)$).

Maximise the object function ; set $\frac{dQ}{d\lambda} = 0$. $\frac{d\theta}{dp_1} = 0$ $\frac{d\theta}{dp_2} = 0$.

we get:

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \frac{\lambda p_1^{h(x_i)} (1-p_1)^{t(x_i)}}{\lambda p_1^{h(x_i)} (1-p_1)^{t(x_i)} + (1-\lambda) p_2^{h(x_i)} (1-p_2)^{t(x_i)}}$$

$$\hat{p}_{1\,new} = \frac{\sum_{i=1}^{N} \frac{\lambda p_1^{h(x_i)} (1-p_1)^{t(x_i)} \cdot h(x_i)}{\lambda p_1^{h(x_i)} + (1-p_1)^{t(x_i)} + (1-\lambda) p_2^{h(x_i)} (1-p_2)^{t(x_i)}}}{M \sum_{i=1}^{N} \frac{\lambda p_1^{h(x_i)} - (1-p_1)^{t(x_i)}}{\lambda p_1^{h(x_i)} + (1-p_1)^{t(x_i)} + (1-\lambda) p_2^{h(x_i)} (1-p_2)^{t(x_i)}}}$$

$$\hat{p}_{2\,new} = \frac{\sum_{i=1}^{N} \frac{\lambda p_2^{h(x_i)} (1-p_2)^{t(x_i)} \cdot h(x_i)}{\lambda p_1^{h(x_i)} (1-p_1)^{t(x_i)} + (1-\lambda) p_2^{h(x_i)} (1-p_2)^{t(x_i)}}}{M \sum_{i=1}^{N} \frac{\lambda p_2^{h(x_i)} (1-p_2)^{t(x_i)}}{\lambda p_1^{h(x_i)} (1-p_1)^{t(x_i)} + (1-\lambda) p_2^{h(x_i)} (1-p_2)^{t(x_i)}}}$$

$$\theta^{old} = (\lambda, p_1, p_2)$$