Exercises for the course

Machine Learning 2

Summer semester 2016

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 6

Exercise 1: RBM with Ternary Hidden Units (40 P)

We consider a variant of the restricted Boltzmann machine where the hidden units $\boldsymbol{h} = (h_1, \dots, h_N)$ are no longer binary, but instead are taking values $\forall_{j=1}^N : h_j \in \{-1, 0, +1\}$. The visible units $\boldsymbol{x} = (x_1, \dots, x_d)$ remain binary with $\forall_{i=1}^d : x_i \in \{0, 1\}$. We consider the usual energy function

$$E_\theta(\boldsymbol{x}, \boldsymbol{h}) = -\boldsymbol{x}^\top W \boldsymbol{h} - \boldsymbol{x}^\top \boldsymbol{a} - \boldsymbol{h}^\top \boldsymbol{b}$$

where $\boldsymbol{a} \in \mathbb{R}^d$, $\boldsymbol{b} \in \mathbb{R}^N$, and $W \in \mathbb{R}^{d \times N}$ are the parameters of the model learned from the data. The energy function maps joint configurations of variables $(\boldsymbol{x}, \boldsymbol{h})$ to probabilities through the probability function

$$p_\theta(\boldsymbol{x}, \boldsymbol{h}) = \frac{1}{Z} e^{-E_\theta(\boldsymbol{x}, \boldsymbol{h})},$$

where Z is the partition function that normalizes the probability distribution to 1. The probability distribution on the visible units is obtained by marginalization of the joint distribution:

$$p_\theta(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in \{-1,0,1\}^N} p(\boldsymbol{x}, \boldsymbol{h}),$$

where we have to sum over 3^N possible configurations of hidden variables.

(a) *Show* that the probability distribution associated to this modified RBM can be rewritten without explicit hidden state space summation as

$$p_\theta(\boldsymbol{x}) = \frac{1}{Z'} e^{-F_\theta(\boldsymbol{x})},$$

with the free energy

$$F_\theta(\boldsymbol{x}) = -\boldsymbol{a}^\top \boldsymbol{x} - \sum_{j=1}^N \log(2 + \cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)),$$

where cosh is the hyperbolic cosine function, Z' normalizes the probability distribution to 1, and $\boldsymbol{w}_j \in \mathbb{R}^d$ is the jth column of the weight matrix W.

(b) *Compute* the partial derivatives of the Kullback-Leibler training objective of the RBM with respect to the parameters a_i, b_j and w_{ij} with $1 \leq i \leq d$ and $1 \leq j \leq N$. That is compute each component of the gradient $\nabla_\theta \,\mathrm{KL}(\hat{p} \,\|\, p_\theta)$ where $\hat{p}$ is the data distribution and p_θ is the distribution modeled by the modified RBM. For this exercise, you can use the result

$$\nabla_\theta \,\mathrm{KL}(\hat{p} \,\|\, p_\theta) = \big\langle \nabla_\theta F_\theta(\boldsymbol{x}) \big\rangle_{\hat{p}} - \big\langle \nabla_\theta F_\theta(\boldsymbol{x}) \big\rangle_{p_\theta}$$

shown during the lecture.

(c) *Compute* the conditional probabilities

$$\Pr(h_j = -1 \mid \boldsymbol{x}) \qquad \Pr(h_j = 0 \mid \boldsymbol{x}) \qquad \Pr(h_j = +1 \mid \boldsymbol{x})$$

$$\Pr(x_i = 0 \mid \boldsymbol{h}) \qquad \Pr(x_i = 1 \mid \boldsymbol{h})$$

required for Gibbs sampling of p_θ associated to the modified RBM.

Exercise 2: Programming Exercise (60 P)

Download the code for Exercise sheet 6 on ISIS and follow the instructions.