Exercises for the course
Machine Learning 1
Winter semester 2020/21

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 13

Exercise 1: RBM with Ternary Hidden Units (20 + 10 P)

We consider a variant of the restricted Boltzmann machine with ternary hidden units $\boldsymbol{h} \in \{-1,0,1\}^H$. Input features remain binary, i.e. $\boldsymbol{x} \in \{0,1\}^d$, like for the classical RBM. The probability model is given by:

$$p(\boldsymbol{x},\boldsymbol{h}|\theta) = \frac{1}{\mathcal{Z}} \exp\Big(\sum_{j=1}^{H} \boldsymbol{w}_j^\top \boldsymbol{x} \cdot h_j + \sum_{j=1}^{H} h_j b_j\Big)$$

where $\theta = (\boldsymbol{w}_j, b_j)_{j=1}^H$ are the parameters of the model, and where and $\mathcal{Z}$ is the partition function that normalizes the probability distribution to 1.

(a) *Show* that this modified RBM can also be expressed as a product of experts

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x},\theta_j),$$

with

$$g_j(\boldsymbol{x},\theta_j) = 1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j),$$

where cosh is the hyperbolic cosine function.

(b) *Show* that the gradient of the log-likelihood assigned to some data point $\boldsymbol{x}_n$ by the modified RBM has the form

$$\forall_{j=1}^H: \quad \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial \boldsymbol{w}_j} = \boldsymbol{x}_n \cdot \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\boldsymbol{x} \cdot \sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$
$$\forall_{j=1}^H: \quad \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial b_j} = \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$

where $\sigma(t) = \dfrac{\sinh(t)}{0.5 + \cosh(t)}$.

Exercise 2: Product of Gaussian Mixture Models (20 + 10 P)

Consider the product of experts:

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x},\theta_j)$$

where each expert is a Gaussian mixture model in d-dimensions, and where each element of the mixture is Gaussian with identity covariance:

$$\forall_{j=1}^H: \; g_j(\boldsymbol{x},\theta_j) = \sum_{k=1}^{C} \alpha_{jk} \frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{\mu}_{jk}\|^2\Big).$$

(a) *Show* that $p(\boldsymbol{x}|\theta)$ can be rewritten as a mixture of C^H elements, where each mixture element (indexed by the vector $\boldsymbol{k} \in \{1,\dots,C\}^H$) has center

$$\boldsymbol{m}_{\boldsymbol{k}} = \frac{1}{H} \sum_{j=1}^{H} \boldsymbol{\mu}_{jk_j}.$$

(b) *Give* the centers $\boldsymbol{m}_{\boldsymbol{k}}$ of the mixture model equivalent to a product of two mixture models, where each mixture model in the product has 2 elements, where the first mixture has the two-dimensional centers $\boldsymbol{\mu}_{11} = \binom{2}{0}$ and $\boldsymbol{\mu}_{12} = \binom{4}{0}$, and where the second mixture has the two-dimensional centers $\boldsymbol{\mu}_{21} = \binom{0}{2}$ and $\boldsymbol{\mu}_{22} = \binom{0}{4}$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.