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 4

Exercise 1: Sparse Coding (5+5 P)

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \mathbb{R}^d$ be a dataset of N examples. Let $\boldsymbol{s}_i \in \mathbb{R}^h$ be the source associated to example $\boldsymbol{x}_i$, and $W \in \mathbb{R}^{d \times h}$ be a matrix of size $d \times h$ that linearly projects the source onto the reconstructed example $\widehat{\boldsymbol{x}}_i$. We optimize the following sparse coding objective:

$$\min_{W, \boldsymbol{s}_1, \dots, \boldsymbol{s}_N} \ \eta \|W\|_F^2 + \sum_{i=1}^N \|\boldsymbol{x}_i - W\boldsymbol{s}_i\|^2 + \lambda \|\boldsymbol{s}_i\|_1 \qquad \text{where} \quad \forall_{i=1}^N : \ \boldsymbol{s}_i \geq 0$$

(a) *Compute* the gradient of the objective with respect to the model parameters W. (i.e. compute the matrix $\frac{\partial E}{\partial W}$).

(b) *Compute* the gradient of the objective with respect to the sources $\boldsymbol{s}_i$ for each data point.

Exercise 2: Sparsifying Non-Linearities (10+10 P)

As an alternative to the sparse coding problem above, we would like to minimize the reparameterized objective of the form:

$$\min_{W, \boldsymbol{r}_1, \dots, \boldsymbol{r}_N} \ \eta \|W\|_F^2 + \sum_{i=1}^N \|\boldsymbol{x}_i - W\boldsymbol{g}(\boldsymbol{r}_i)\|^2 + \lambda \|\boldsymbol{r}_i\|^2 \qquad \text{where} \quad \forall_{i=1}^N : \ \boldsymbol{r}_i \in \mathbb{R}^h$$

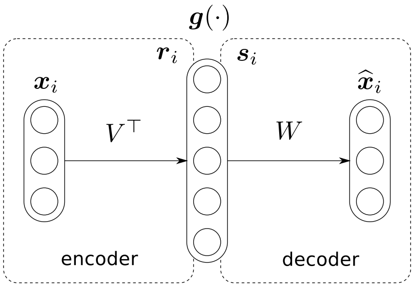
We call $\boldsymbol{r}_i$ the source parameter, and $\boldsymbol{s}_i = \boldsymbol{g}(\boldsymbol{r}_i)$ the reparameterized source associated to example $\boldsymbol{x}_i$. Note that the new objective no longer involves the minimization of an L_1-norm, and also does not include positivity constraints.

(a) *Find* a reparameterization function $\boldsymbol{g} : \mathbb{R}^h \to \mathbb{R}^h$ for which the optimization problem is equivalent to the one of Exercise 1.

(b) *Explain* what are the advantages and disadvantages of using such formulation of the optimization problem when compared to the original sparse coding problem. Your answer may include: (1) Applicability of gradient descent to find sources $\boldsymbol{s}_i$. (2) Ease of using an encoder to initialize the search for optimal sources.

Exercise 3: Auto-Encoders (10+10 P)

We now give an explicit definition of the encoder $\boldsymbol{r}_i = V^\top \boldsymbol{x}_i$, where $V \in \mathbb{R}^{d \times h}$ is a matrix of size $d \times h$. A graphical depiction of the resulting auto-encoder for $\boldsymbol{x}_i \in \mathbb{R}^3$ and $\boldsymbol{r}_i, \boldsymbol{s}_i \in \mathbb{R}^5$ is given below:



(a) Assuming the same error function as in Exercise 2, *use* the chain rule to express the gradient of the objective with respect to the encoder parameter $\frac{\partial E}{\partial V}$.

(b) *Explain* what are the advantages and disadvantages of using an autoencoder instead of directly optimizing $\boldsymbol{s}_i$ or $\boldsymbol{r}_i$. Your answer should include the following aspects: (1) Computational requirements of inferring sources $\boldsymbol{s}_i$ from observations $\boldsymbol{x}_i$. (2) Difficulty of the optimization problem. (3) Computational requirements at training time.

Exercise 4: Programming Exercise (50 P)

Download the code for Exercise sheet 4 on ISIS and follow the instructions.