TOPICS IN STATISTICAL SCIENCES II – EXAM EXERCISE 4

SØREN HØJSGAARD AND CHRISTIAN HIRSCH

This exercise is about deep learning. During the oral exam you will have 20 min to present the exercise. You decide what topics to cover and how to present them, however, we will ask questions to any part of the covered curricula, exercise and presentation.

1. Variational Autoencoder

Recall that in the context of variational autoencoders, the evidence lower bound is defined via $\mathbb{E}_q[\log(p(x|z))] - \mathsf{KL}(q|p)$, where $\mathsf{KL}(q|p)$ denotes the Kullback-Leibler distance between q and p.

(1) Show that the evidence lower bound equals $\mathbb{E}_q[\log(p(x,z))] - \mathbb{E}_q[\log(q(z))]$.
(2) Show that the evidence lower bound is indeed smaller than $\log(p(x))$.
(3) Suppose that $p(x,z) = p(x,z;\theta)$ depends on a parameter θ. Choose some initial θ^0 and then iteratively define the θ^{i+1} by optimizing the ELBO with the choice $q(z) = p(z|x;\theta^i)$. Show that this is equivalent to the EM algorithm. What prevents us from choosing $q(z) = p(z|x;\theta^i)$ in practice?
(4) Let p be the density of the multivariate standard normal distribution in $\mathbb{R}^n$ and let q be the density of the normal distribution with mean vector $\boldsymbol{\mu} = (\mu_i)_{i\le n}$ and a diagonal covariance matrix with diagonal entries $\boldsymbol{\sigma^2} = (\sigma_i^2)_{i\le n}$. Show that

$$\mathsf{KL}(q|p) = -\frac{1}{2}\sum_{i\le n}(\log(\sigma_i^2) - \mu_i^2 - \sigma_i^2) + c,$$

where $c \in \mathbb{R}$ does not depend on $\boldsymbol{\mu}$ or $\boldsymbol{\sigma^2}$.

2. Sun Spot Prediction

The `R` dataset `sunspot.month` describes the number of sunspots since 1749. Analyze this time series via deep learning. The goal is to predict the value of the current month from the 50 previous ones.

(1) How do you prepare the data such that the problem falls in the framework of supervised learning? Also think of a sound separation into training and validation data.
(2) How would an architecture for a suitable recurrent neural network look like?
(3) How would an architecture for a suitable convolutional neural network look like?
(4) How can you combine a convolutional and a recurrent neural network?
(5) How do the models compare in terms of training and validation error?
(6) How do the models compare in training and prediction speed? Explain your observations.
(7) Provide visualizations on how the predicted values compare to the true ones.