Stochastic Machine Learning
Chapter 04 - Deep Kalman Filters

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2024

Deep Kalman Filters

- Deep Kalman filters have been proposed in Krishnan, Shalit, and Sontag (2015) and we want to shortly revisit this work

Example (Kalman filter)

- The unobserved signal $X = (X_n)_{n \geq 1}$ is a Gaussian process given by
$$X_n = aX_{n-1} + b\epsilon_n,$$
where (ϵ_n) are i.i.d. $\mathcal{N}(0, 1)$.
- The observation is given by
$$Y_n = AX_n + B\eta_n,$$
where also (η_n) are also i.i.d., independent of (ϵ_n).
- Hence, the parameter vector θ is given by $(a, b, A, B, \mu_0, \Sigma_0)$.

- The goal of this work is to replace the linear relationships by neural networks.
- Start with
$$X_0 \sim \mathcal{N}(\mu_0, \Sigma_0).$$
- Assume that
$$X_n \sim \mathcal{N}\Big(NN^1(X_{n-1}, \epsilon_n), NN^2(X_{n-1}, \epsilon_n)\Big),$$
i.e. a probabilisitic model where mean and standard deviation are parametrized through a neural network
- The observation
$$Y_n = F\big(NN^3(X_n)\big),$$
where we have a parametric family of distributions F_θ with $\theta \in \Theta$ at hand.
- However, the log-likelihood is difficult to maximize.
- As in the EM-algorithm, the authors aim at a suitable lower bound.
- They show that the log-likelihood can be approximated by some Kullback-Leibler distance (using as above Jensen's inequality).
- The Kullback-Leibler distance for normal distributions is easy to compute and they arrive at a differential loss function which is then implemented.
- Also numerical results are given.

Krishnan, Rahul G, Uri Shalit, and David Sontag (2015). „Deep kalman filters“. In: **Proc. Conf. Neural Inf. Process. Syst. Workshops**.