Machine Learning with Signal Processing
Part IV: Application Examples

Arno Solin
Assistant Professor in Machine Learning
Department of Computer Science
Aalto University

ICML 2020 Tutorial

@arnosolin
arno.solin.fi
Outline

- Non-Gaussian likelihoods
- Spatio-temporal GPs
- Latent-space GPs
- Latent differential equations
- Langevin dynamics
- Summary
Gaussian processes ♥ SDEs

GPs under the kernel formalism

\[ f(t) \sim \text{GP}(0, \kappa(t, t')) \]
\[ y \mid f \sim \prod_i p(y_i \mid f(t_i)) \]

Stochastic differential equations

\[ df(t) = F f(t) + L d\beta(t) \]
\[ y_i \sim p(y_i \mid h^T f(t_i)) \]

Flexible model specification
Inference / First-principles
Non-Gaussian likelihoods

- The observation model might not be Gaussian

\[ f(t) \sim \text{GP}(0, \kappa(t, t')) \]
\[ y \mid f \sim \prod_i \rho(y_i \mid f(t_i)) \]

- There exists a multitude of great methods to tackle general likelihoods with approximations of the form

\[ \mathbb{Q}(f \mid \mathcal{D}) = \mathcal{N}(f \mid m + K\alpha, (K^{-1} + W)^{-1}) \]

- Use those methods, but deal with the latent using state space models
Inference

- Laplace approximation
- Variational Bayes
- Direct KL minimization
- EP or Assumed density filtering (Single-sweep EP)
- Can be evaluated in terms of a (Kalman) filter forward and backward pass, or by iterating them
Example

- Commercial aircraft accidents 1919–2017
- Log-Gaussian Cox process (Poisson likelihood) by ADF/EP
- Daily binning, $n = 35,959$
- GP prior with a covariance function:

$$\kappa(t, t') = \kappa_{\text{Mat.}}^{\nu=3/2} (t, t') + \kappa_{\text{Per.}}^{\text{year}} (t, t') \kappa_{\text{Mat.}}^{\nu=3/2} (t, t') + \kappa_{\text{Per.}}^{\text{week}} (t, t') \kappa_{\text{Mat.}}^{\nu=3/2} (t, t')$$

- Learn hyperparameters by optimizing the marginal likelihood

Nickisch et al. (2018). ICML.
Example

Commercial aircraft accidents 1919–2017

Log-Gaussian Cox process (Poisson likelihood) by ADF/EP

Daily binning, \( n = 35, 959 \)

GP prior with a covariance function:

\[
\kappa(t, t') = \kappa_{\text{Mat.}}(t, t') + \kappa_{\text{year}}(t, t') + \kappa_{\text{week}}(t, t')
\]

Learn hyperparameters by optimizing the marginal likelihood

Nickisch et al. (2018). ICML.
Example

Commercial aircraft accidents 1919–2017

Log-Gaussian Cox process (Poisson likelihood) by ADF/EP

Daily binning, \( n = 35, 959 \)

GP prior with a covariance function:

\[
\kappa(t, t') = \kappa_{\text{Mat.}}(t, t') + \kappa_{\text{year}} + \kappa_{\text{week}}
\]

Learn hyperparameters by optimizing the marginal likelihood

Nickisch et al. (2018). ICML.
Spatio-temporal GPs

\[ f(r) \sim \text{GP}(0, \kappa(r, r')) \]
\[ y \mid f \sim \prod_i p(y_i \mid f(r_i)) \]

\[ f(x, t) \sim \text{GP}(0, \kappa(x, t; x', t')) \]
\[ y \mid f \sim \prod_i p(y_i \mid f(x_i, t_i)) \]
Spatio-temporal Gaussian processes

GPs under the kernel formalism

\[ f(x, t) \sim \text{GP}(0, k(x, t; x', t')) \]
\[ y_i = f(x_i, t_i) + \varepsilon_i \]

Stochastic partial differential equations

\[ \frac{\partial f(x, t)}{\partial t} = \mathcal{F} f(x, t) + \mathcal{L} w(x, t) \]
\[ y_i = \mathcal{H}_i f(x, t) + \varepsilon_i \]
Spatio-temporal GP regression

Spatio-temporal GP regression

Spatio-temporal GP classification

Wilkinson et al. (2020). ICML.
Multi-view stereo as a temporal fusion problem

▶ Inputs: Frame pairs and relative camera poses

▶ Encoder–decoder network for depth estimation

▶ Treat the encoder as a feature extractor, and do GP regression in the latent space

▶ The GP prior encodes the similarity of the camera views

Hou et al. (2019). ICCV.
Hou et al. (2019). ICCV. Video: https://youtu.be/ie1lGr1NW7k
Sequential priors in GAN latent space

Hou et al. Submitted.
Latent differential equations

- Define an **implicit prior** over functions through dynamics
- Define **observation likelihoods**. Anything differentiable w.r.t. latent state (e.g. text models!)
- Train everything jointly with automatic differentiation + variational inference

Adopted form David Duvenaud.
Latent ODEs

- Learning low-rank latent representations of possibly high-dimensional sequential data trajectories
- Combination of *variational auto-encoders (VAEs)* with sequential data with a latent space governed by a *continuous-time ODE*

Yildiz et al. (2019). NeurIPS.
In SDEs: Need to store noise

- Infinite reparameterization trick: Use same Brownian motion sample on forward and reverse passes

Li et al. (2020). AISTATS.
Latent SDE models

Ornstein–Uhlenbeck prior
Laplace likelihood

MOCAP example
50D data, 6D latent space,
11000 params

Li et al. (2020). AISTATS.
Langevin dynamics

- Molecular simulation
- Sampling and parameter inference (e.g., Metropolis-adjusted Langevin algorithm)
- Diffusion models (even for images!)

Langevin’s model for brownian motion

Denoising diffusion probabilistic model with Langevin dynamics

Ho et al. Submitted.
Summary

Part I
Tools and discrete-time models

Part II
SDEs (continuous-time models)

Part III
Gaussian processes

Part IV
Application examples
Acknowledgements

Simo Särkkä (Aalto)
Manon Kok (TU Delft)
Thomas Schön (Uppsala)
Richard Turner (Cambridge)

James Hensman (Amazon)
Nicolas Durrande (prowler.io)
Michael Riis Andersen (DTU)
Emtiyaz Khan (RIKEN)

The following researchers kindly provided material/examples:
David Duvenaud, Markus Heinonen, Cagatay Yildiz, Jonathan Ho
Acknowledgements

My group at Aalto University
The examples in this tutorial were related to the following publications:


References
The examples in this tutorial were related to the following publications:


Machine learning with signal processing

Part I
Tools and discrete-time models

Part II
SDEs (continuous-time models)

Part III
Gaussian processes

Part IV
Application examples