DS-GA 3001.001 Special Topics in Data Science: Probabilistic Time Series Analysis
Homework 4

Due date: Nov 22

Problem 1. (15pt) Which of these objects are a Gaussian process?

- linear combination of 2 GPs: $f(x) = af_1(x) + bf_2(x)$ where $f_i \sim \mathcal{GP}(\mu_i(x); k_i(x, y))$ (independent) and a, b are fixed parameters.

- random linear: $f(x) = ax + w$ where $a \sim \mathcal{N}(0, \sigma_a^2)$, $w \sim \mathcal{N}(0, \sigma_w^2)$.

- random periodic: $f(x) = a\cos(wx) + b\sin(wx)$ with $a \sim \mathcal{N}(0, \sigma^2)$, $b \sim \mathcal{N}(0, \sigma^2)$, w fixed parameter.

If yes, then write down the corresponding mean and covariance functions.

Problem 2. (10pt) How would you construct a GP-equivalent of an ARIMA (1,2,1) model?

Problem 3. (15pt) Derive the mean and covariance of $P(y|\theta)$ for the FITC approximation described in the lecture (this is obtained by marginalizing out $\mathbf{u}$ and $\mathbf{f}$).
Hint: one can think of the approximate model as a sequence of linear gaussian steps and use the usual simple gaussid.pdf properties.

Problem 4. (10pt) What GP-based model would you use for the Johnson&Johnson quarterly earnings database? Explain your choices. Would it matter if the goal of your analysis is to interpolate to account for missing data in the middle of the recorded time interval vs. extrapolating a decade into the future?