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

Due date: April 26th, by midnight

Problem 1. (10pt) Derive the mean and covariance of $\mathrm{P}(y_t|\theta)$ for the FITC approximation.
Hint: think of the approximate model as a sequence of linear gaussian steps.

Problem 2. (10pt) Consider a simple GP regression model with gaussian observations: $\mathbf{f}\ \mathcal{GP}(0, k(\cdot,\cdot))$, $y_i \sim \mathcal{N}(f_i, \sigma^2)$. Let $\mathrm{Var}(f(x^*))_n$ be the variance of the posterior for the value of function f evaluated at a test position x^*, conditioned on n observations $x_{1:n}, y_{1:n}$. If $\mathrm{Var}(f(x^*))_{n-1}$ denotes the same variance conditioned on the first $n-1$ observations, show that the posterior variance decreases by adding the last datapoint, i.e. $\mathrm{Var}(f(x^*))_n \leq \mathrm{Var}(f(x^*))_{n-1}$.
Hint: This is problem 4 from section 2.9 GP textbook (pg.31). See detailed suggestions there.

Problem 3. extra credit (10pt) Fit a GP to the Johnson&Johnson quarterly earnings dataset. Chose 2 kernel variants and briefly justify your choices. Using the remaining data as training set, predict the values of the function in the period 1970-1975 and 1980-1990 (extrapolation in the future). Compare the two models by comparing the average prediction error at the test points 1970-1975 and 1980-1985. Comment on results.