

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

**Due date: Nov 22**  
**YG390**

**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. The distribution of the sum of the functions due to linearity of Gaussian distributions is simply another  $\mathcal{GP}$ :  $f(x) \sim \mathcal{GP}(\mu_1(x) + \mu_2(x); k_1(x, y) + k_2(x, y))$
- random linear:  $f(x) = ax + w$  where  $a \sim \mathcal{N}(0, \sigma_a^2)$ ,  $w \sim \mathcal{N}(0, \sigma_w^2)$ . We consider two fixed  $x_1$  and  $x_2$  and note that the two random variables  $f(x_1)$  and  $f(x_2)$  are jointly Gaussian.

$$\begin{aligned} \mathbb{E}[f(x_i)] &= \mathbb{E}[ax_i + w] \\ &= x_i \mathbb{E}[a] + \mathbb{E}[w] \\ &= 0 \\ \text{Var}[f(x)] &= \text{Var}[ax + w] \\ &= x^2 \text{Var}[a] + \text{Var}[w] \\ &= x^2 \sigma_a^2 + \sigma_w^2 \\ \text{Covar}[f(x_1), f(x_2)] &= \mathbb{E}(ax_1 + w)(ax_2 + w) \\ &= x_1 x_2 \mathbb{E}[a, a] + \mathbb{E}[w, w] \\ &= x_1 x_2 \sigma_a^2 + \sigma_w^2 \end{aligned}$$

Let  $y_1 = f(x_1)$  and  $y_2 = f(x_2)$ , the multivariate gaussian distribution  $p(y_1, y_2)$  is:

$$p(y_1, y_2) = \mathcal{N}\left(\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} x_1^2 \sigma_a^2 + \sigma_w^2 & x_1 x_2 \sigma_a^2 + \sigma_w^2 \\ x_1 x_2 \sigma_a^2 + \sigma_w^2 & x_2^2 \sigma_a^2 + \sigma_w^2 \end{bmatrix}\right)$$

Thus  $f(x) \sim \mathcal{GP}(\mu(x_1, x_2); \Sigma(x_1, x_2))$ .

- 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. Similarly, we have now:

$$\begin{aligned} \mathbb{E}[f(x_i)] &= \mathbb{E}[a \cos(wx_i) + b \sin(wx_i)] \\ &= \cos(wx_i) \mathbb{E}[a] + \sin(wx_i) \mathbb{E}[b] \\ &= 0 \\ \text{Var}[f(x)] &= \text{Var}[a \cos(wx) + b \sin(wx)] \\ &= \cos^2(wx) \text{Var}[a] + \sin^2(wx) \text{Var}[b] \\ &= \sigma^2 \\ \text{Covar}[f(x_1), f(x_2)] &= \mathbb{E}[(a \cos(wx_1) + b \sin(wx_1))(a \cos(wx_2) + b \sin(wx_2))] \\ &= \cos(wx_1) \cos(wx_2) \mathbb{E}[a, a] + \sin(wx_1) \sin(wx_2) \mathbb{E}[b, b] \\ &= \cos(wx_1) \cos(wx_2) \sigma^2 + \sin(wx_1) \sin(wx_2) \sigma^2 \\ &= \sigma^2 \cos(w(x_1 - x_2)) \end{aligned}$$

$f(x)$  is a Gaussian process:  $f(x) \sim \mathcal{GP}(\mu(x_1, x_2); \Sigma(x_1, x_2))$ , where  $\mu(x_1, x_2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  and  $\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2 \cos(w(x_1 - x_2)) \\ \sigma^2 \cos(w(x_1 - x_2)) & \sigma^2 \end{bmatrix}$

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.*

Given the two factor graphs in the slides of the lecture,  $\{y_i\}$  are conditionally independent given  $\{f_i\}$  and depend directly on the  $\{f_i\}$  which are conditionally independent given  $\{u_i\}$ . The  $\mathbf{u}$  variables summarize the dependencies in  $\mathbf{f}$ . Setting  $p(\mathbf{u}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{uu})$ , the conditional distribution  $p(f_t|\mathbf{u}) = \mathcal{N}(f_t; K_{f_t u} K_{uu}^{-1} \mathbf{u}, K_{f_t f_t} - K_{f_t u} K_{uu}^{-1} K_{u f_t})$ .

We have then:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{f}_t + \sigma_y \boldsymbol{\epsilon} \text{ with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ \text{let } D &= K_{f_t f_t} - K_{f_t u} K_{uu}^{-1} K_{u f_t} \\ \mathbf{f}_t &= K_{f_t u} K_{uu}^{-1} \mathbf{u}_t + D \boldsymbol{\epsilon}' \text{ with } \boldsymbol{\epsilon}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \end{aligned}$$

$\mathbf{f}_t$  has 0 mean as it only depends on  $\mathbf{u}_t$  which has 0 mean, the same for  $\mathbf{y}_t$  which depends only on  $\mathbf{f}_t$  with some noise. Now:

$$\begin{aligned} \text{Covar}[\mathbf{y}_t, \mathbf{y}_s] &= \mathbb{E}_{\mathbf{u}, \mathbf{f}, \boldsymbol{\epsilon}} [(\mathbf{f}_t + \sigma_y \boldsymbol{\epsilon})(\mathbf{f}_s + \sigma_y \boldsymbol{\epsilon})^T] \\ &= \mathbb{E}[\mathbf{f}_t, \mathbf{f}_s] + \sigma_y^2 \mathbb{E}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] \\ &= \mathbb{E}[\mathbf{f}_t, \mathbf{f}_s] + \sigma_y^2 \mathbf{I} \\ \mathbb{E}[\mathbf{f}_t, \mathbf{f}_s] &= \mathbb{E}[K_{f_t u} K_{uu}^{-1} \mathbf{u}_t \mathbf{u}_s^T K_{u f_s}] + D \mathbb{E}[\boldsymbol{\epsilon}', \boldsymbol{\epsilon}'] \\ &= K_{f_t u} K_{uu}^{-1} \mathbb{E}[\mathbf{u}_t, \mathbf{u}_s] K_{u f_s}^T + D \mathbf{I} \\ &= K_{f_t u} K_{uu}^{-1} K_{u u} K_{u f_s} + D \\ &= K_{f_t u} K_{uu}^{-1} K_{u f_s} + D \end{aligned}$$

Where the cross-terms in the second and fourth equalities disappear as these terms are mutually independent. Thus

$$P(\mathbf{y}|\theta) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{fu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uf} + \mathbf{D} + \sigma_y^2 \mathbf{I}).$$

**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?

I used a linear combination of squared exponential kernel, linear, polynomial and periodic kernels. I tested the effect of the number of observations (sampling of the data) for the interpolation and forecasting (see 1). Experimenting with different weighting of the kernels, I found that a mixed of SE and periodic kernels, captured the best the increasing trend, and some of the wiggles of the time series (see 2). The error bars showed that when the observations were sparse the GP model confidence interval increased very rapidly confirming that GP was less confident in its predictions. The same combination of kernels for the forecasting task, for different horizons, had a harder time to predict the wiggles of the process in the future (see 3). Note the difference between interpolation when the GP uses neighboring points vs. extrapolating in the future when the only available points are the last observations before the forecasting (backcast). So for interpolating to account for missing data, the best choice could be an SE + linear or polynomial model which takes into account clusters of points and for forecasting an Exp-Sine-Squared or spectral mixture kernels might be more appropriate in an effort to give weight to close observations, but at same time captures the dynamics of the process. In order to improve the overall model performance, I used a spectral mixture kernel which seemed to predict the swings of the data including the ones in the future (see 4).

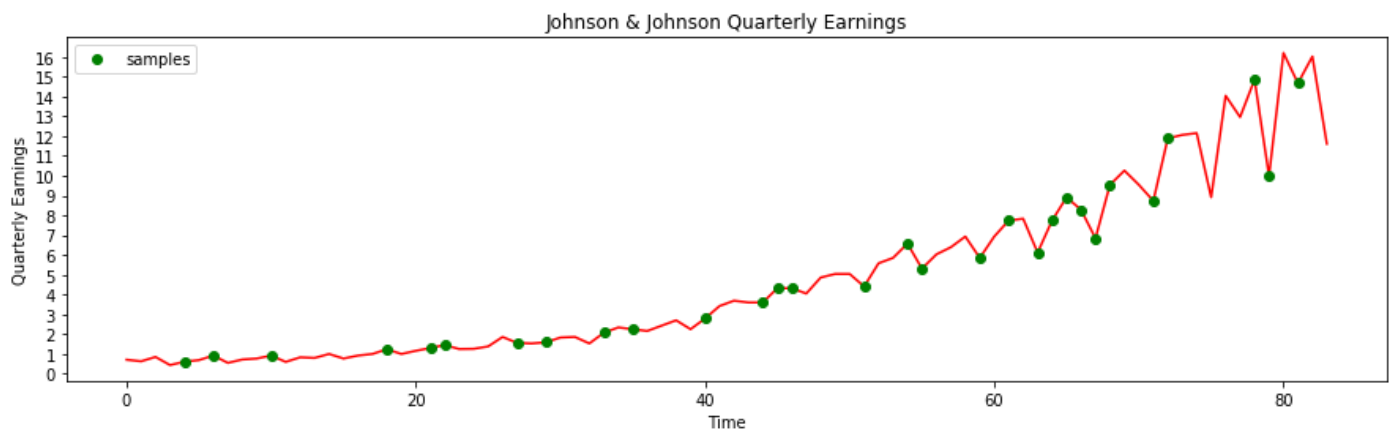


Figure 1: Johnson&Johnson

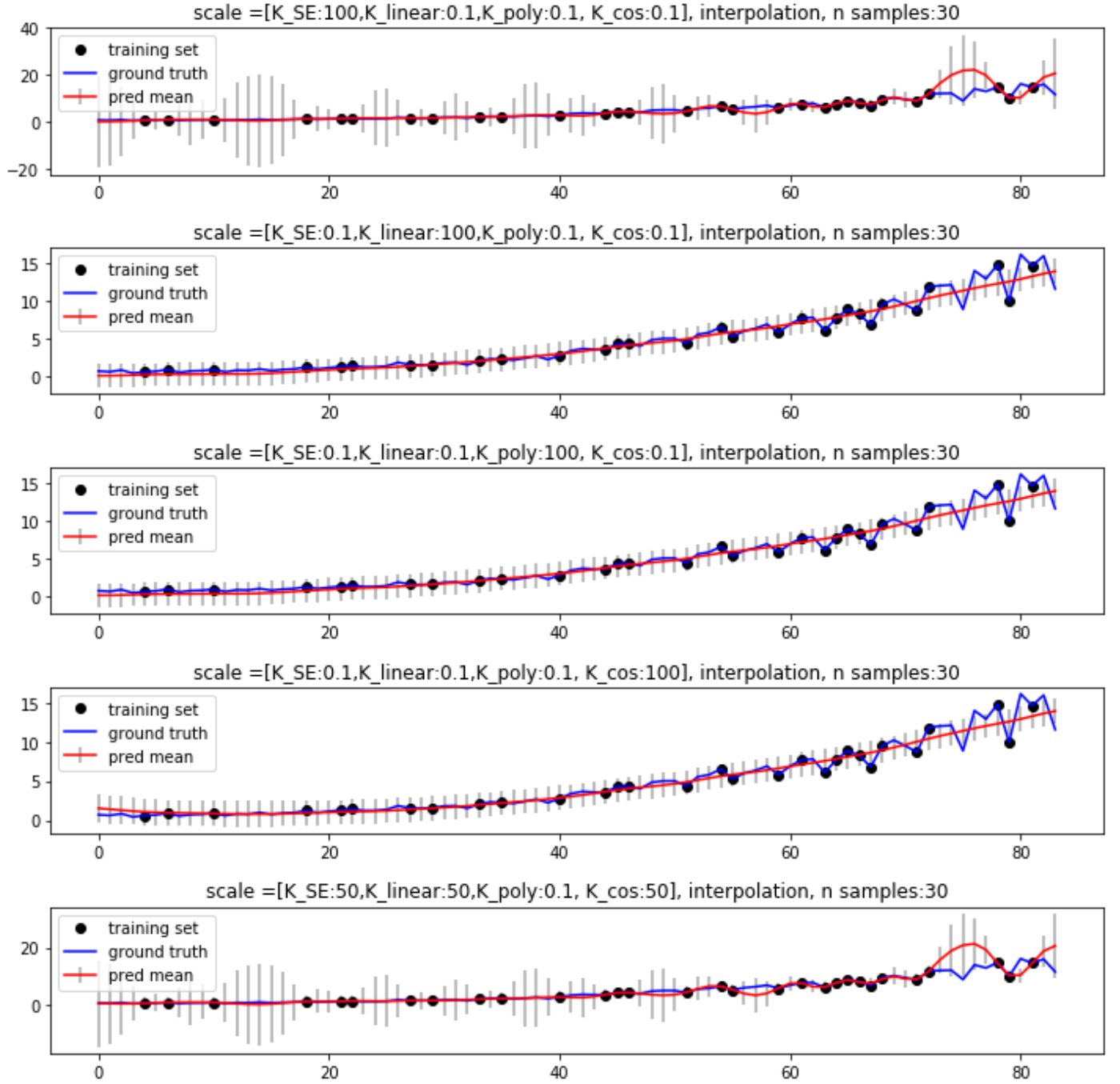


Figure 2: Interpolation using a weighted combinations of kernels

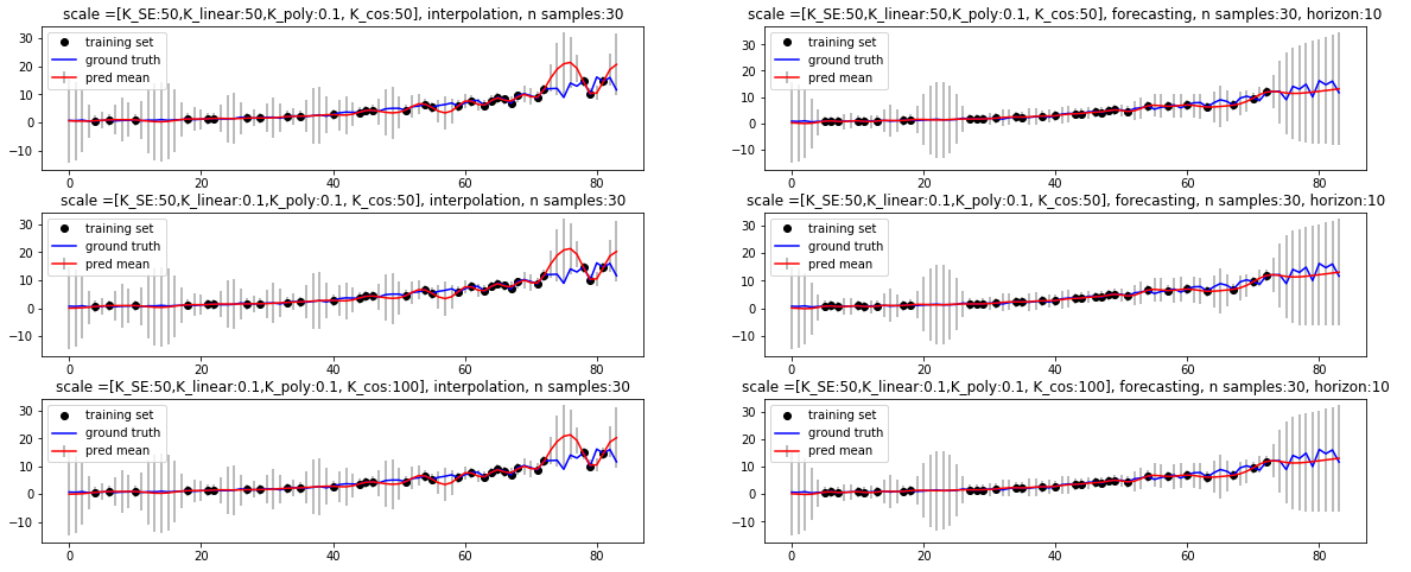


Figure 3: SE + periodic kernels - Interpolation and Forecasting with error bars

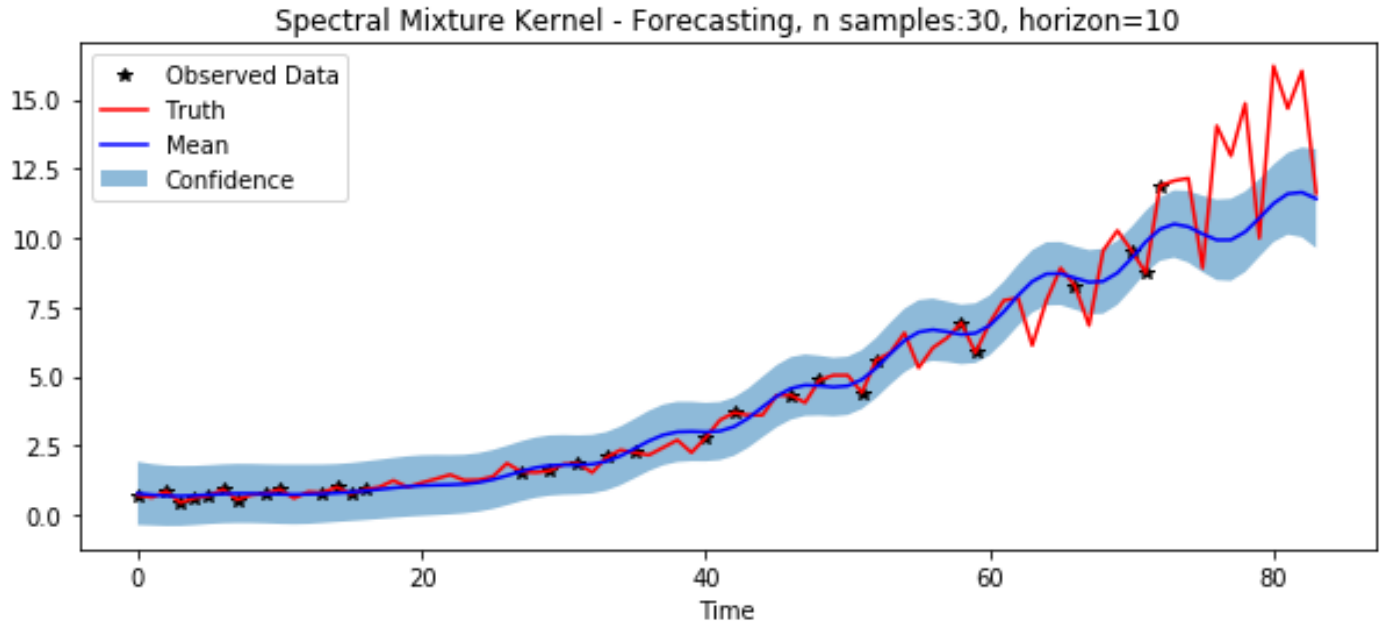


Figure 4: Spectral Mixture Kernel with confidence intervals - two standard deviations