Probabilistic Inference & Learning

Exercise Sheet #5

Gaussian Process Regression

1. **Modeling Body Weight.** Download the code and data from the lecture (`WeightTrackModel.ipynb` and `weightdata_clean.mat`) and address the following questions.

   (a) Say I was planning to pick up running again on 1 August 2013. Assuming that I did so (and managed to keep at it), while not doing any of the other specific activities tracked in the model, what is the prediction (mean $\pm$ standard-deviation) for my bodyweight (relative to 0 on 4 March 2009, the first datum in the dataset) for 31 Dec 2013? (Use the features and model as presented in the lecture)

   30 points

   (b) Given the model as used in the lecture, what is the *joint* posterior over the 5 parametric feature weights given the data? Given an intuitive explanation for the covariance structure of this posterior (you may find it helpful to consider the plot produced in the cell titled "plot of the features" in the notebook). *20 points*

2. **Derivatives of Gaussian Process Samples.** For this exercise, consider the random function $f : \mathbb{R} \to \mathbb{R}$ from the centered (i.e. zero-mean) GP prior $p(f) = \mathcal{GP}(f, 0, k)$ with the "square-exponential" kernel

$$k(a,b) = \theta^2 \exp\left(-\frac{(a-b)^2}{2\lambda^2}\right) \quad \text{with parameters } \theta, \lambda \in \mathbb{R}_+$$

   (a) given n locations $X = [x_1, x_2, \dots, x_n]$, what is the marginal prior distribution over $f_X := [f(x_1), f(x_2), \dots, f(x_n)]$? *5 points*

   (b) Consider arbitrary linear maps (i.e. matrices) $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{t\times n}$. What is the marginal prior distribution over $f_A := Af_X$? Assume we have access to given observations $Y = [y_1, \dots, y_m]$ generated according to the likelihood

$$p(Y \mid f_X) = \mathcal{N}(Y; Af_X, \sigma^2 I_m)$$

   What is the *posterior* distribution $p(f_B \mid Y)$ over the values $f_B := Bf_X$? *10 points*

   (c) Now consider the *derivative* $f'(x) = {}^{d}/_{dx} f(x)$. What is the prior distribution $p(f')$ implied by $p(f) = \mathcal{GP}(f; 0, k)$. To find the answer, notice that the operator ${}^{d}/_{dx}$ is a *linear* operator (i.e. ${}^{d}/_{dx}(\alpha f_1(x) + \beta f_2(x)) = \alpha f_1'(x) + \beta f_2'(x)$). Thus, we can think of it as an 'infinitesimal' matrix acting on f (if you find this mathematically problematic, just suspend disbelief for the purposes of this exercise and treat the operator like the matrix A in the question above, and f_X as a potentially infinitely long vector). Show that the marginal is itself a Gaussian process. What is the kernel (covariance function) of this GP over f'? *15 points*

   (d) Consider the observations $Y = [y_1, y_2]$ with likelihood (notice the derivative at x_2!)

$$p(Y \mid f, f') = \mathcal{N}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \begin{bmatrix} f(x_1) \\ f'(x_2) \end{bmatrix}, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)$$

   What is the posterior over $f(x)$ arising from these observations? *20 points*

   For your own intuition, although not necessary to gain full points, it may be helpful to make a plot, using suitable values for all parameters in this problem. E.g., make plots over $x \in [-3, 3]$ and consider

$$\theta = \lambda = 1, \sigma = 0.1, x_1 = -1, x_2 = 1, y_1 = 1, y_2 = -1.$$