Exercise sheet: Gaussian processes

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time to solve compared to an exercise with (***), which is a more complex exercise.

1. (*) Let $f(t) = \int_0^t u(\tau)d\tau$. If $u(t) \sim \mathcal{GP}(0, k_u(t,t'))$, i.e. $u(t)$ is a GP with kernel function $k_u(t,t')$, write the expression that corresponds to the kernel function for $f(t)$, i.e. $k_f(t,t')$.

2. (*) The linear kernel is defined as $k(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top \mathbf{z}$. If $\mathbf{X}$ is a design matrix of input vectors,
$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix},$$
write the expression for the kernel matrix $\mathbf{K}$ in terms of the matrix $\mathbf{X}$.

3. (**) Using the properties for the marginal and conditional Gaussians (see Appendix A below) show that the posterior distribution for $p(\mathbf{w}|\mathbf{y}, \mathbf{X})$ is given as
$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = \mathcal{N}(\mathbf{w}|\frac{1}{\sigma_n^2}\mathbf{A}^{-1}\mathbf{\Phi}^\top\mathbf{y}, \mathbf{A}^{-1}),$$
where $\mathbf{A} = \sigma_n^{-2}\mathbf{\Phi}^\top\mathbf{\Phi} + \mathbf{\Sigma}_p^{-1}$, with $\mathbf{\Phi} \in \mathbb{R}^{n\times N}$.

4. (*) Show that the predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y})$ is given as
$$p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}\left(f_* \middle| \frac{1}{\sigma_n^2}\boldsymbol{\phi}(\mathbf{x}_*)^\top\mathbf{A}^{-1}\mathbf{\Phi}^\top\mathbf{y}, \boldsymbol{\phi}(\mathbf{x}_*)^\top\mathbf{A}^{-1}\boldsymbol{\phi}(\mathbf{x}_*)\right),$$
where $\mathbf{A} = \sigma_n^{-2}\mathbf{\Phi}^\top\mathbf{\Phi} + \mathbf{\Sigma}_p^{-1}$.

5. (**) Show that another way to write the predictive distribution from the previous exercise is
$$\begin{aligned} p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}\Big(f_* \Big| & \boldsymbol{\phi}_*^\top\mathbf{\Sigma}_p\mathbf{\Phi}^\top(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\mathbf{y}, \\ & \boldsymbol{\phi}_*^\top\mathbf{\Sigma}_p\boldsymbol{\phi}_* - \boldsymbol{\phi}_*^\top\mathbf{\Sigma}_p\mathbf{\Phi}^\top(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\mathbf{\Phi}\mathbf{\Sigma}_p\boldsymbol{\phi}_*\Big), \end{aligned}$$
where $\boldsymbol{\phi}(\mathbf{x}_*) = \boldsymbol{\phi}_*$, y $\mathbf{K} = \mathbf{\Phi}\mathbf{\Sigma}_p\mathbf{\Phi}^\top$.
[HINT: use the properties for the matrix inverses shown in Appendix B]

6. (*) Show that if $k_1(\mathbf{x}, \mathbf{x}')$ is a valid kernel, then $k(\mathbf{x}, \mathbf{x}') = ck_1(\mathbf{x}, \mathbf{x}')$, with $c > 0$ is a valid kernel.

7. (*) Show that $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{A} \mathbf{x}'$ is a valid kernel, with $\mathbf{A}$ a symmetric positive semidefinite matrix.

8. (**) Let $\text{var}_n(f(\mathbf{x}_*))$ be the predictive variance of a Gaussian process regression model at $\mathbf{x}_*$ given a dataset of size n. The corresponding predictive variance using a dataset of only the first $n-1$ training points is denoted $\text{var}_{n-1}(f(\mathbf{x}_*))$. Show that $\text{var}_n(f(\mathbf{x}_*)) \leq \text{var}_{n-1}(f(\mathbf{x}_*))$, i.e. that the predictive variance at $\mathbf{x}_*$ cannot increase as more training data is obtained.
[HINT: use the inverse of a partitioned matrix as shown in Appendix B]

Appendix A: marginal and conditional Gaussians

Given a marginal Gaussian distribution for $\mathbf{x}$, and a conditional Gaussian distribution for $\mathbf{y}$ given $\mathbf{x}$,

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{B}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}),$$

the marginal distribution for $\mathbf{y}$, and the conditional distribution for $\mathbf{x}$ given $\mathbf{y}$ are given by

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{B}\boldsymbol{\Lambda}^{-1}\mathbf{B}^\top)$$
$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{B}^\top\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{B}^\top\mathbf{L}\mathbf{B})^{-1}.$$

Appendix B: matrix identities involving inverses

A useful identity involving matrix inverses is the following

$$\left(\mathbf{P}^{-1} + \mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^\top\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^\top\left(\mathbf{B}\mathbf{P}\mathbf{B}^\top + \mathbf{R}\right)^{-1}.$$

Say $\mathbf{P} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{M\times M}$, so that $\mathbf{B} \in \mathbb{R}^{M\times N}$. If $M \ll N$, it is much cheaper to evaluate the right-hand side of the expression above than the left-hand side.

Another useful identity involving inverses is the following:

$$\left(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1},$$

which is known as the *Woodbury identity.* This is useful, for instance, when $\mathbf{A}$ is large and diagonal, and hence easy to invert, while $\mathbf{B}$ has many rows but few columns (and conversely for $\mathbf{C}$) so that the right-hand side is much cheaper to evaluate than the left-hand side.

One more useful identity involving inverses is the following. Let the invertible $n \times n$ matrix $\mathbf{A}$ and its inverse $\mathbf{A}^{-1}$ be partitioned into

$$\mathbf{A} = \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{pmatrix}, \quad \mathbf{A}^{-1} = \begin{pmatrix} \tilde{\mathbf{P}} & \tilde{\mathbf{Q}} \\ \tilde{\mathbf{R}} & \tilde{\mathbf{S}} \end{pmatrix},$$

where $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are $n_1 \times n_1$ matrices and $\mathbf{S}$ and $\tilde{\mathbf{S}}$ are $n_2 \times n_2$ matrices with $n = n_1 + n_2$. The submatrices of $\mathbf{A}^{-1}$ are given

$$\left.\begin{array}{ll} \tilde{\mathbf{P}} & = \mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{Q}\mathbf{M}\mathbf{R}\mathbf{P}^{-1} \\ \tilde{\mathbf{Q}} & = -\mathbf{P}^{-1}\mathbf{Q}\mathbf{M} \\ \tilde{\mathbf{R}} & = -\mathbf{M}\mathbf{R}\mathbf{P}^{-1} \\ \tilde{\mathbf{S}} & = \mathbf{M} \end{array}\right\} \text{ where } \mathbf{M} = \left(\mathbf{S} - \mathbf{R}\mathbf{P}^{-1}\mathbf{Q}\right)^{-1}$$