Lab 2 - On Gaussian Processes

F. Septier

1 Exercise 1 - Sampling from a Gaussian process prior

In this exercise, we will write the code needed to draw and plot samples of $f(\cdot)$ from a Gaussian process prior with a squared exponential (or, equivalently, RBF) kernel, more specifically $f : \mathbb{R} \mapsto \mathbb{R}$ such that

$$f(\cdot) \sim \mathcal{GP}(m(\cdot), \kappa(\cdot, \cdot)) \text{ with } m(x) = 0 \text{ and } \kappa(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}||x - x'||^2\right)$$

To implement this, we choose a vector of m test input points $\boldsymbol{x}^* = \begin{bmatrix} x_1{}^* & \dots & x_m^* \end{bmatrix}^T$. We will choose $\boldsymbol{x}^*$ to contain sufficiently many points, such that it will appear as a continuous function on the screen. We then evaluate the $m \times m$ covariance matrix

$$K(\boldsymbol{x}^*, \boldsymbol{x}^*) = \begin{bmatrix} \kappa(x_1^*, x_1^*) & \cdots & \kappa(x_1^*, x_m^*) \\ \vdots & & \vdots \\ \kappa(x_m^*, x_1^*) & \cdots & \kappa(x_m^*, x_m^*) \end{bmatrix}$$

and thereafter generate samples from the multivariate normal distribution

$$f(\boldsymbol{x}^*) \sim \mathcal{N}\left(m(\boldsymbol{x}^*), K(\boldsymbol{x}^*, \boldsymbol{x}^*)\right)$$

(a) Use `numpy.linspace` to construct a vector $\boldsymbol{x}^*$ with $m = 101$ elements equally spaced between $[-5; 5]$

(b) Construct a mean vector $m(\boldsymbol{x}^*)$ with 101 elements all equal to zero and the 101×101 covariance matrix $K(\boldsymbol{x}^*, \boldsymbol{x}^*)$. The expression for $\kappa(\cdot, \cdot)$ is given above. Let the hyperparameters be $l = 2$ and $\sigma_f^2 = 1$.
Hint : You might find it useful to define a function that returns $\kappa(x, x')$, taking x and x' as arguments.
Hint : One solution to construct the matrix rapidly using numpy is to use for example the computation of the distance as
`numpy.abs(xs[:,numpy.newaxis]-xs[:,numpy.newaxis].T)**2)`
where `xs` is the grid $\boldsymbol{x}^*$, i.e. the output of the `numpy.linspace`.

(c) Use `scipy.stats.multivariate_normal` (you might need to use the option `allow_singular=True`) to draw 25 realizations of the function on the 101 grid points, i.e. $f^{(1)}(\boldsymbol{x}^*), \dots, f^{(25)}(\boldsymbol{x}^*)$ sample independently from the multivariate normal distribution $f(\boldsymbol{x}^*) \sim \mathcal{N}\left(m(\boldsymbol{x}^*), K(\boldsymbol{x}^*, \boldsymbol{x}^*)\right)$

(d) Plot these samples $f^{(1)}(\boldsymbol{x}^*), \dots, f^{(25)}(\boldsymbol{x}^*)$ versus the input vector $\boldsymbol{x}^*$

(e) Try another value of l and repeat steps (b)-(d). How do the two plots differ and why ?

2 Exercise 2 - Posterior of the Gaussian process

In this exercise we will perform Gaussian process regression as seen in the lecture. That means, based on the N (noiseless) observations $\mathcal{D} = \left\{x^{(i)}, y = f(x^{(i)})\right\}_{i=1}^N$ and the prior belief $f(\cdot) \sim \mathcal{GP}(m(\cdot), \kappa(\cdot, \cdot))$, we want to find the posterior $p(f|\mathcal{D})$. (In the previous problem, we were only concerned with the prior $p(f)$, not conditioned on having observed the data $\mathcal{D}$.) We consider the same Gaussian process prior (same mean $m(x)$, covariance kernel $\kappa(x, x')$ and hyperparameters) as in the previous exercise.

(a) Construct two vectors $\boldsymbol{x} = \begin{bmatrix}-4 & -3 & -1 & 0 & 2\end{bmatrix}^T$ and $\boldsymbol{x} = \begin{bmatrix}-2 & 0 & 1 & 2 & -1\end{bmatrix}^T$ which will be our training data (that is, $N = 5$).

(b) Keep $\boldsymbol{x}^*$ as in the previous problem. In addition to the $m \times m$ matrix $K(\boldsymbol{x}^*, \boldsymbol{x}^*)$, now also compute the $N \times m$ matrix $K(\boldsymbol{x}, \boldsymbol{x}^*)$ and the $N \times N$ matrix $K(\boldsymbol{x}, \boldsymbol{x})$.

(c) Use the training data $(\boldsymbol{x}, \boldsymbol{y})$ and the matrices constructed in (b) to compute the posterior mean $\boldsymbol{\mu}_{\text{post}}$ and the posterior covariance $\boldsymbol{\Sigma}_{\text{post}}$ for $\boldsymbol{x}^*$, by using the equations for conditional multivariate normal distributions.

(d) In a similar manner as in (c) and (d) in the previous problem, draw 25 samples from the multivariate distribution $f(\boldsymbol{x}^*)|\mathcal{D} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$ and plot these samples ($f^{(j)}(\boldsymbol{x}^*)$ vs. $\boldsymbol{x}^*$) together with the posterior mean ($\boldsymbol{\mu}_{\text{post}}$ vs. $\boldsymbol{x}^*$) and the actual measurements ($f(\boldsymbol{x}^*)$ vs $\boldsymbol{x}^*$). How do the samples in this plot differ from the prior samples in the previous problem ?

(e) Instead of plotting samples, plot a credibility region. Here, a credibility region is based on the (marginal) posterior variance. The 68% credibility region, for example, is the area between $\boldsymbol{\mu}_{\text{post}} - \sqrt{\text{diag}(\boldsymbol{\Sigma}_{\text{post}})}$ and $\boldsymbol{\mu}_{\text{post}} - \sqrt{\text{diag}(\boldsymbol{\Sigma}_{\text{post}})}$, where $\text{diag}(\boldsymbol{\Sigma}_{\text{post}})$ is a vector with the diagonal elements of $\boldsymbol{\Sigma}_{\text{post}}$. What is the connection between the credibility regions and the samples you drew previously ?
Hint : the python command `matplotlib.pyplot.fill_between` could be used.

(f) Now, consider the setting where the measurements are corrupted with noise, $y_i = f(x_i) + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Use $\sigma = 0.1$ and repeat (c)-(e) with this modification of the model. What is the difference in comparison to the previous plot ? What is the interpretation ?

(g) Explore what happens with another length scale l.

(h) The squared exponential kernel/covariance function gives samples which are smooth and infinitely continuously differentiable. Other kernels make other assumptions. Now try the previous problems using the exponential kernel instead,

$$\kappa(x, x') = \exp\left(-\frac{1}{l}|x - x'|\right)$$