

Machine Learning for Earth System Sciences Assignment 1

Due date: 12th February 2022

[All submissions are to be done independently. This will be assessed based on the values chosen in steps 1 and 2 below. Submissions containing identical values will be considered to have been plagiarized, and penalty will be imposed accordingly. There are no fixed answers to the problems given below. The approaches taken will be assessed for grading.]

A1. Spatial Process

Consider a square region of size 20x20 and 100 time-steps.

- 1) At $t=1$, select 10 random locations (likely to be different for all students). Put down observed values $X(s,t)$ according to the formula $X(s,t) = s_1 + s_2 + \text{random noise}$, where s_1, s_2 are the horizontal and vertical positions of the location s in the square.
- 2) Using this initial data, use a Gaussian Process to generate data for the whole region (fitting the initial data) and for all the time-points.

Reference for fitting GP on samples:

https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_prior_posterior.html

Use different kinds of Gaussian Process kernels as discussed here: https://scikit-learn.org/stable/modules/gaussian_process.html

- 3) Once the data has been generated, pick up a random set of 20 locations. Consider that you know the values at these locations for all time-points. Use Kriging to estimate the values at other locations. Compare the estimated values with generated values.

Kriging reference: http://users.ictp.it/~pub_off/lectures/lns018/06Barua2.pdf

Python implementation: <https://geostat-framework.readthedocs.io/projects/pykrige/en/stable/>

A2. Spatio-temporal process

Repeat the initialization of step 1 of A1, at 5 random time-points and 5 spatial locations (25 points overall), setting $X(s,t) = s_1 + s_2 + 0.5 \cdot t + \text{random noise}$. Repeat step 2 using spatio-temporally separable Gaussian Processes (i.e. decompose as a product of two GPs – one spatial and one temporal). Explain how the generated dataset is different from the dataset obtained in Step 2.