Machine Learning For Earth System Sciences
Assignment 1

by

Abhranil Chandra (19MF10002)

Department of Mechanical Engineering

Indian Institute of Technology Kharagpur

An Assignment Report Submitted for the
Fulfilment of the Machine Learning for Earth System Sciences (AI60002)
Course Requirements

Professor: **Prof. Adway Mitra**

Problem Statement

1. **Spatial Process**: Consider a square region of size 20 x 20 and 100 time-steps.

   (a) 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 $X(s, t) = s_1 + s_2 + \epsilon$, where s_1, s_2 are the horizontal and vertical positions of the location s in the square and ϵ is the random noise.

   (b) 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.

   (c) 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.

2. **Spatio-temporal process**: Repeat the initialization of step 1 of Spatial Process, at 5 random time-points and 5 spatial locations (25 points overall), setting $X(s, t) = s_1 + s_2 + 0.5t + \epsilon$. 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.

General Introduction

Gaussian Process

Consider the training data: $D = \{ f(x_i, y_i) \mid i = 1, 2, ..., n\}$ .Each input is a vector x_i of dimension d and each target is a real-valued scalar $y_i = f(x_i)$ for some unknown function f, possibly corrupted by noise. Let us construct the design matrix $X \in R_{d \times n}$ and the outputs vector $y \in R^n$ resulting in the input-output training set $D = \{X, y\}$.

Assumption: There exists an underlying process for the unknown function f.

Objective: Predict the output y at the input test x i.e., $y = f(x)$ using the training data set D. We can restate the objective as inferring a distribution over function f given the training data i.e., $p(f \mid D = \{X, y\})$, then use this information to make a prediction on y.

A solution to the above problem is a Gaussian Process which is a Bayesian Inference Model.

1. Gaussian Process is a way of defining distribution over functions. It defines a prior over function f and then based on Bayesian inference converts it into a posterior once some data is observed.

2. Gaussian Process is a stochastic process whose realizations consist of a collection of random variables associated with every point in the range of either time or space.

3. In such a process each random variable has a normal distribution. A Gaussian Process is completely specified by a mean function and a positive definite covariance function. This property facilitates model fitting as only the first and second-order moments of the process are required to be specified.

Kriging Model

Kriging is a spatial interpolation method based on the statistical model. Common types of Kriging are as follows.

1. **Simple Kriging**: Simple kriging can be seen as the mean and envelope of Brownian random walks passing through the data points.

2. **Ordinary Kriging**: Ordinary Kriging is a spatial estimation method where the error variance is minimized.

3. **Kriging with a trend**: Kriging with a trend provides the least-squares estimate of the attribute of interest concurrently with the least-squares fit of a previously established deterministic trend function.

4. **Factorial Kriging**: Factorial kriging is a disaggregation that enables the identification of the scale of information and thus provides a means for matching the scale between the two variables.

5. **Co-Kriging**: A Co-Kriging method is a method that is used to predict the value of the point at unobserved locations by sample points are known to be spatially interconnected by adding other variables that have a correlation with the main variable or can also be used to predict 2 or more variables simultaneously.

6. **Kriging based on Gaussian Process modeling**: This is a method of interpolation based on the Gaussian process governed by prior covariances. Under suitable assumptions of the priors, kriging gives the best linear unbiased prediction at unsampled locations.

Solution Approach(Part 1- Spatial Process Only)

A)

We have a square region of size 20 x 20. First, we select 10 random locations inside the square region at $t = 1$. The Spatio-temporal data at these locations are according to the function $X(s, t) = s_1 + s_2 + \epsilon$, where s_1, s_2 are the horizontal and vertical positions with respect to the region and ϵ is random noise which I have taken as the standard normal(normally distributed with zero mean and variance of 10). As we can see from the expression, there is no time dependence for the target variable.

The data thus obtained are enlisted in Table 1. Using this data for $t = 1$, a Gaussian Process was fit in order to get the mean and variance for each location for each time point.

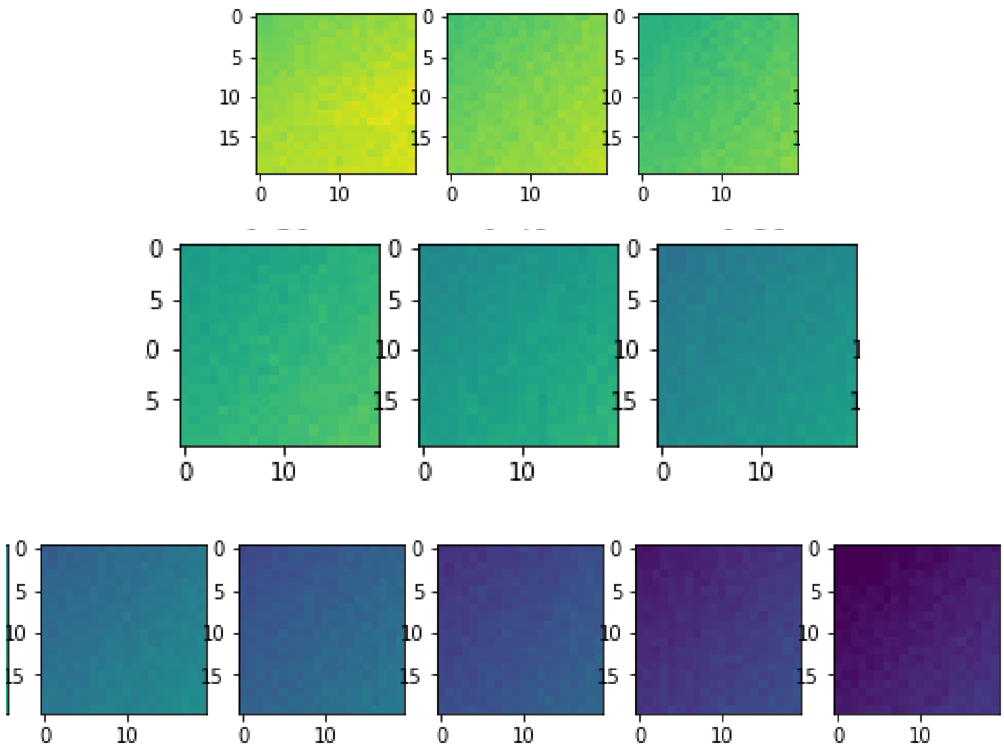
Now, a random set of 20 locations were chosen for $t = 1$. The data for these 20 locations were sampled from the Gaussian Process fit on the previously obtained dataset on 10 locations at $t = 1$. The sampling process is as follows:

Sl No.	X-coordinate	Y-coordinate	Time	St-data
1	3	15	1.000	18.520
2	8	12	1.000	20.730
3	12	0	1.000	12.102
4	9	9	1.000	18.753
5	2	19	1.000	21.709
6	14	4	1.000	18.316
7	13	8	1.000	21.874
8	16	17	1.000	33.048
9	10	1	1.000	11.422
10	7	2	1.000	9.773

Table 1: Spatial data on which Gaussian Process is fitted

B)

Below I show have shown the plots of the point values generated for a location across 10 time slices by fitting Gaussian Process on the above 10 points.



C)

1. The Gaussian Process model is given input with the horizontal position(s_1) and the vertical position(s_2) of the location.
2. The Gaussian Process model returns the mean and variance of the probability distribution (Gaussian) over the functions.
3. Using a normal distribution with the obtained mean and variance, a spatio-temporal value is sampled.

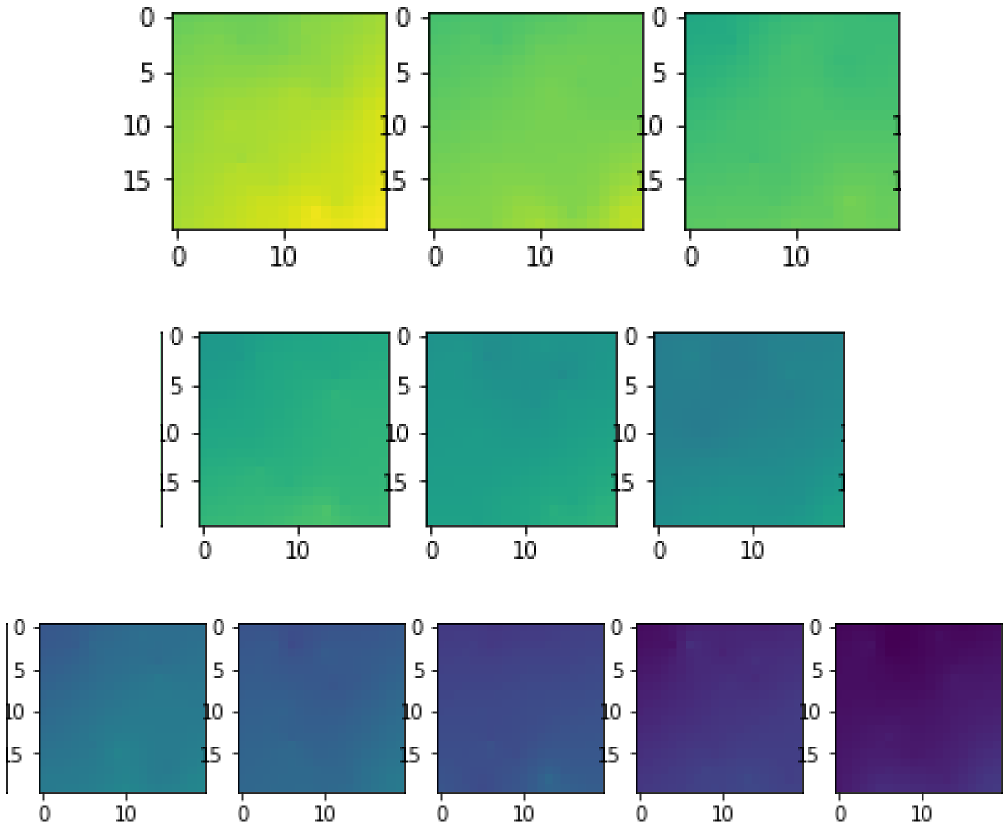
After the dataset was sampled, a Kriging model was used to fit on the sampled dataset.

The data(20 random points at time step 1) on which the Kriging model was fit is:

np.array([np.array([4, 2]), np.array([18, 18]), np.array([0, 19]), np.array([14, 4]), np.array([15, 0]), np.array([12, 1]), np.array([6, 13]), np.array([9, 15]),np.array([6, 14]), np.array([19, 9]),np.array([10, 19]), np.array([15, 17]), np.array([8, 17]), np.array([5, 10]), np.array([13, 18]), np.array([14, 6]), np.array([5, 19]), np.array([11, 7]), np.array([6, 2]), np.array([10, 3])])

The actual target data for these points was obtained from the Gaussian process fitted earlier on the 10 points in table 1.

These are the plots of point values estimated for a location across 10 time slices by fitting the Kriging model on the above random 20 points.



We can clearly see the difference in the values generated by the Gaussian Process and the estimated values by teh Kriging model by seeing teh difference in colours of the contour plots. This is very obvious as the two models learn slightly different underlying function estimation for f , used to estimate these point values.

Solution Approach(Part 2- Spacio-Temporal Process)

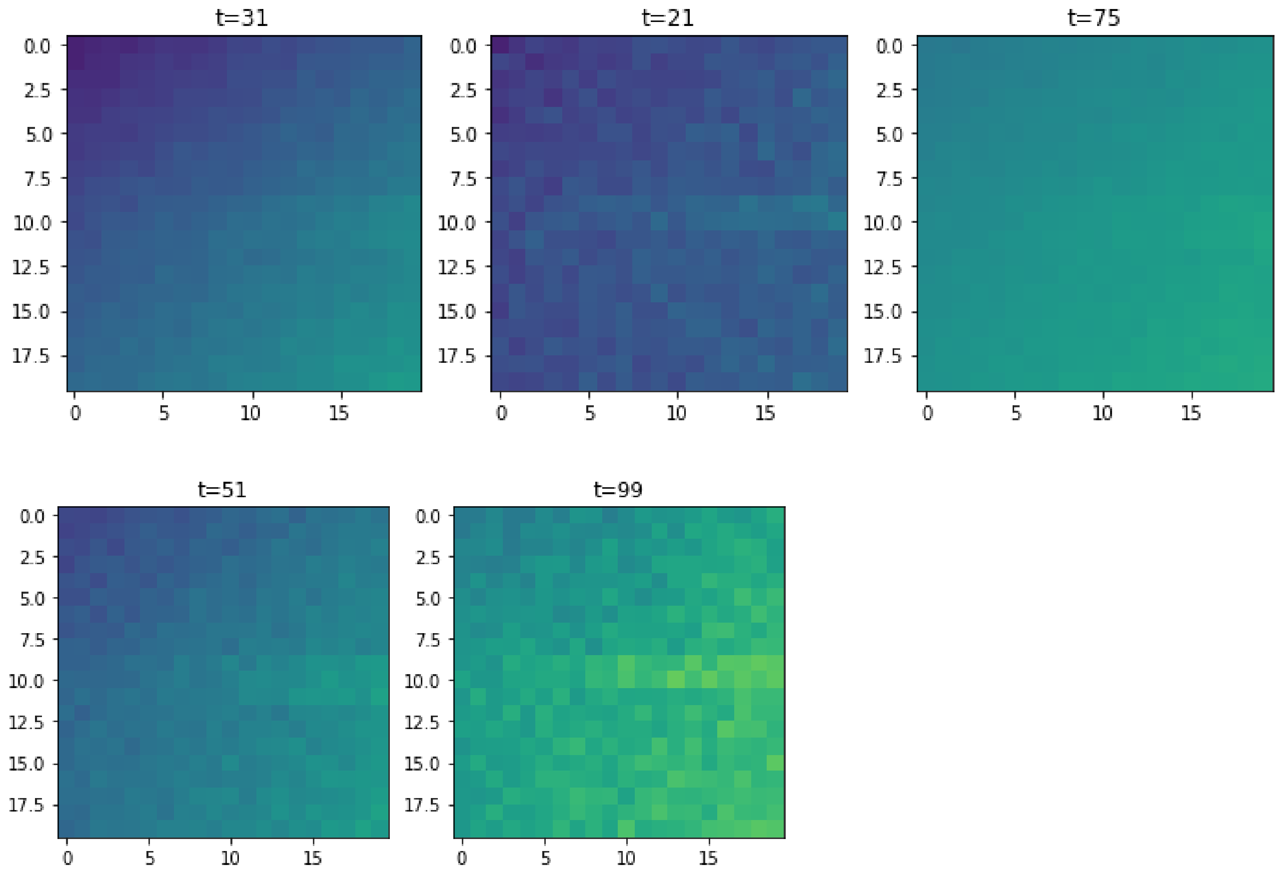
Taking 5 random spatial locations and 5 random temporal points

(25 points overall)

spatial_locs = [np.array([13, 8]), np.array([7, 4]), np.array([10, 19]), np.array([15, 2]), np.array([1, 11])]

temporal_locs = np.array([31, 21, 75, 51, 99])

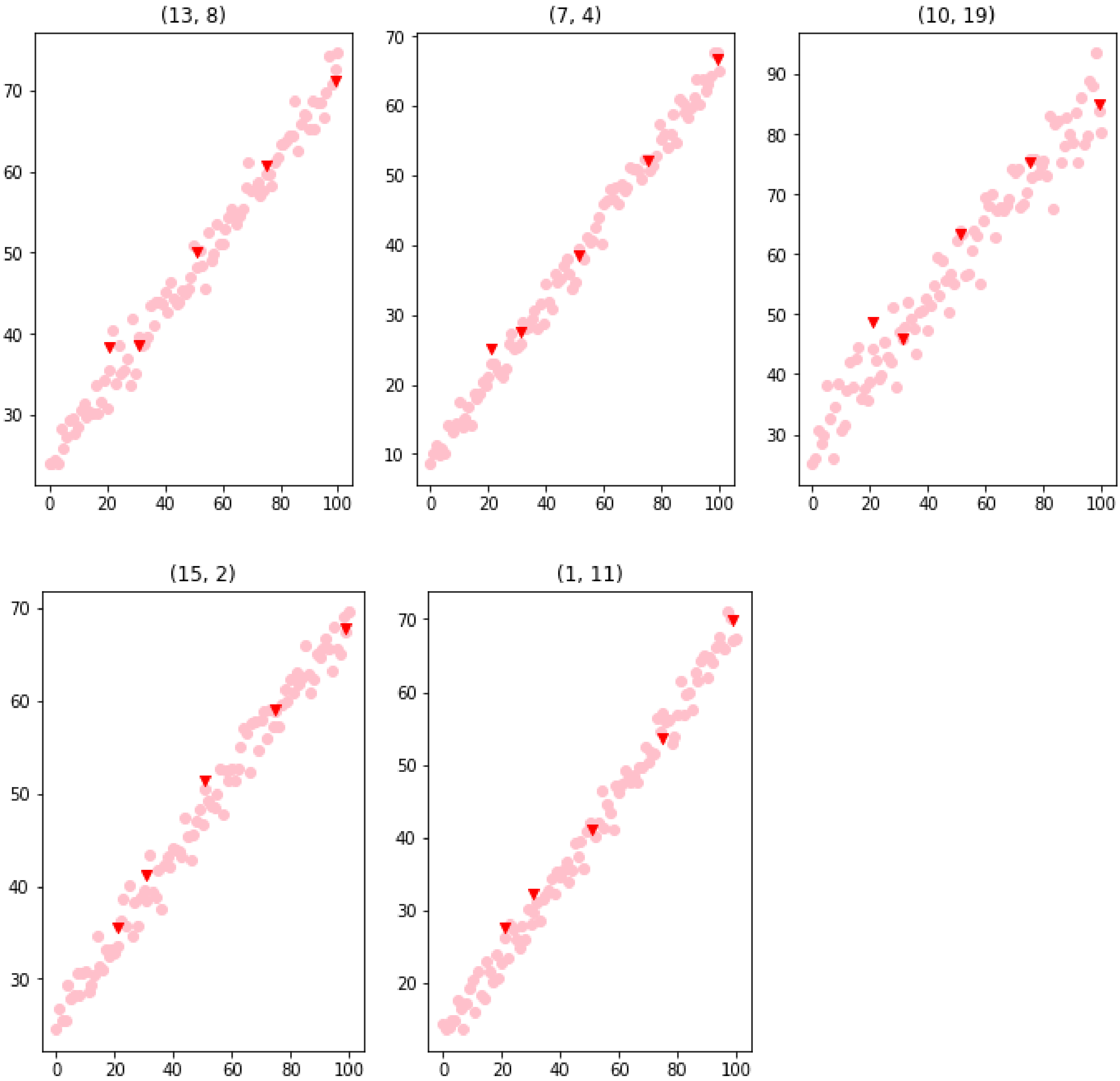
Data fit by Gaussian Process.



We use this fitted Gaussian Process model here using spatio-temporally separable Gaussian Processes (i.e. decompose as a product of two GPs – one spatial and one temporal) to estimate the values of these 5 spatial points over all 100 time-steps.

Then we use the Kriging model to estimate the same as in part 1 and we plot them by superposing over each other to compare the estimated values between the Gaussian process and the Kriging model

Here I plot the values estimated by the Kriging model(in Dark Red) over the fitted Gaussian Process(in Light Red) models over the 25 spacio-temporal data points in the previous page



Final Discussion and Future Work

Finally with respect to the difference between data generated using only spatial data and data generated using both spatial and time data, we can infer that the underlying structure becomes more complex when the time dependence of target is introduced. Also, intuitively, we can say that as the number of features increases, the covariance stationarity of the data may become redundant.

Kriging generates a smoother graph, hence the Gaussian set of graphs are noticeably bumpier. Both of them follow the same general values with minimal differences in plotted points. The graphs generated by the Kriging kernels are smoother than the Gaussian graphs due to the algorithms.

The spatial values are now increasing with time instead of decreasing. This is because there is now information about the change of the graph with time, which was missing from the first part.

We can use different kernels in the models to see their corresponding changes and perform an ablation study. Moreover, we can use other Bayesian Inference techniques to fit the data and do inference on the new data.