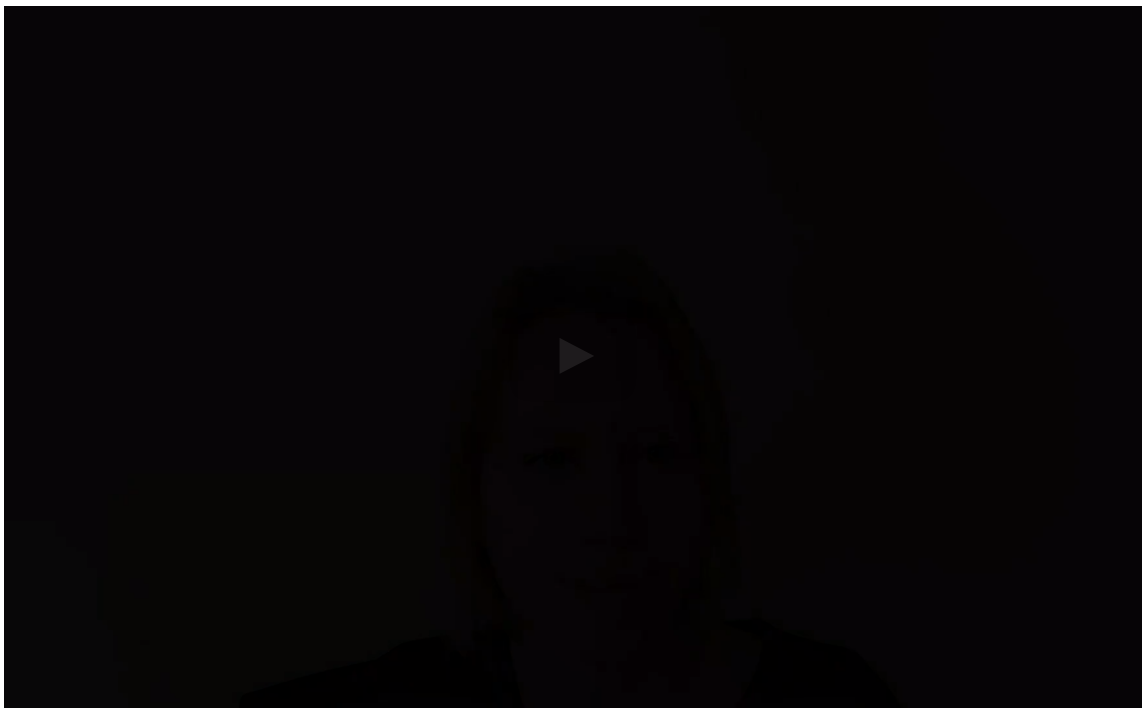
3. How to Model Covariance Matrices

Bookmark this page

Exercises due May 21, 2021 19:59 EDT

How to Model Covariance Matrices



20:45 / 20:45 1.50x

methods we have been talking about earlier in this course.

So now, in summary, again, I have two interpretations.

One is that I just get my prediction

as a weighted average of the measurement

values in the neighborhood weighted

by a function of the distance.

Or I could view this as, essentially,

a linear regression with nonlinear features.

And the features are given by this kernel matrix.

Both of these views are valuable,

and we will see both of these having

effects in what is going to come next in our lecture here.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Video note: At 17:42, α should be referred to as a column vector.

In all previous contents we have assumed we have the covariance matrix available. However, we will often be in a situation where we have made some observations, $\mathbf{x}_2$, but do not have the covariances for these observations. We can instead make an assumption on the covariances.

For example, going back to the case of *City 1* and *City 2*, one can assume that the covariance between the two cities is determined by its distance. In this case,

$$\text{Cov}\,(X_1, X_2) = k\,(Z_1, Z_2)\,,$$

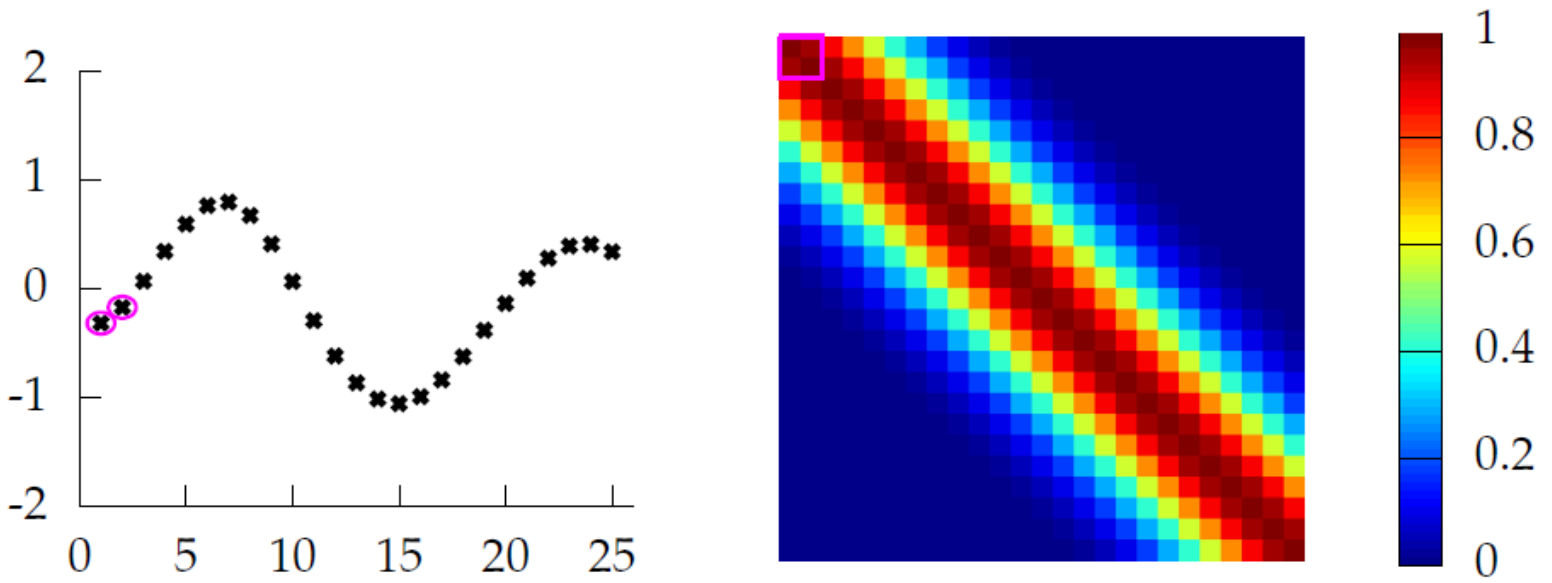
where $k\,(\cdot, \cdot)$ is some covariance function $k\,(Z_1, Z_2)$, for example,

$$k\,(Z_1, Z_2) = \exp\left(-\frac{\|Z_1 - Z_2\|^2}{2\ell^2}\right), \tag{7.5}$$

where ℓ is a parameter to be estimated, called the **length-scale** . Covariance functions are often go by the name *kernels* as well; however, kernels are a broader class of functions and not all kernels are covariance functions.

Recall that Z_1 and Z_2 are points in some space. In our case, Z_1 and Z_2 are the physical locations of *City 1* and *City 2*. Thus, $||Z_1 - Z_2||^2$ measures the Euclidean distance in this space between the two cities.

The figure below shows an example of generating a covariance matrix from a covariance function. The matrix on the right is the generated covariance matrix. The plot on the left is a random draw from a multivariate Normal distribution using this covariance matrix.



(a) A 25 dimensional correlated random variable (values ploted against index)

(b) colormap showing correlations between dimensions.

33: An example of a covariance matrix generated by 25 Normal random variables located at equal intervals over the real line.

We can now ask the question of what happens if we take some of these points on the left-hand plot, and, instead of drawing them from the multivariate Normal distribution, we instead set them to to our observed measurements and condition the remaining points on those set values. We will do this by first partitioning the vector space into the $N - d$ observed values, $\mathbf{x}_2$, and the d unobserved values, $\mathbf{x}_1$, such that the mean becomes

$$\mu_{\mathbf{X}} = \begin{bmatrix} \mu_1 & \in \mathbb{R}^d \\ \mu_2 & \in \mathbb{R}^{N-d} \end{bmatrix}.$$

Note that although we have placed the means for the observed values at the end of this vector, they can be located anywhere on the above plot. The order of the values in the mean vector does not need to correspond to the spatial or temporal order for the values.

We then compute the full covariance matrix using the covariance function:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} \in \mathbb{R}^{d \times d} & \mathbf{\Sigma}_{12} \in \mathbb{R}^{d \times (N-d)} \\ \mathbf{\Sigma}_{21} \in \mathbb{R}^{(N-d) \times d} & \mathbf{\Sigma}_{22} \in \mathbb{R}^{(N-d) \times (N-d)} \end{bmatrix}$$

Finally, we condition the means and covariances for the unobserved values on the observed values:

$$\begin{aligned} \mu_{\mathbf{X}_1|\mathbf{X}_2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) \\ \Sigma_{\mathbf{X}_1|\mathbf{X}_2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

You will notice that to do this, we need some values for μ_1 and μ_2, but we only have the observations $\mathbf{x}_2$. Thus, we need to make an assumption on the means. A common (but not necessary) assumption is that $\mu_{\mathbf{X}} = 0$ so that both mean vectors are also zero.

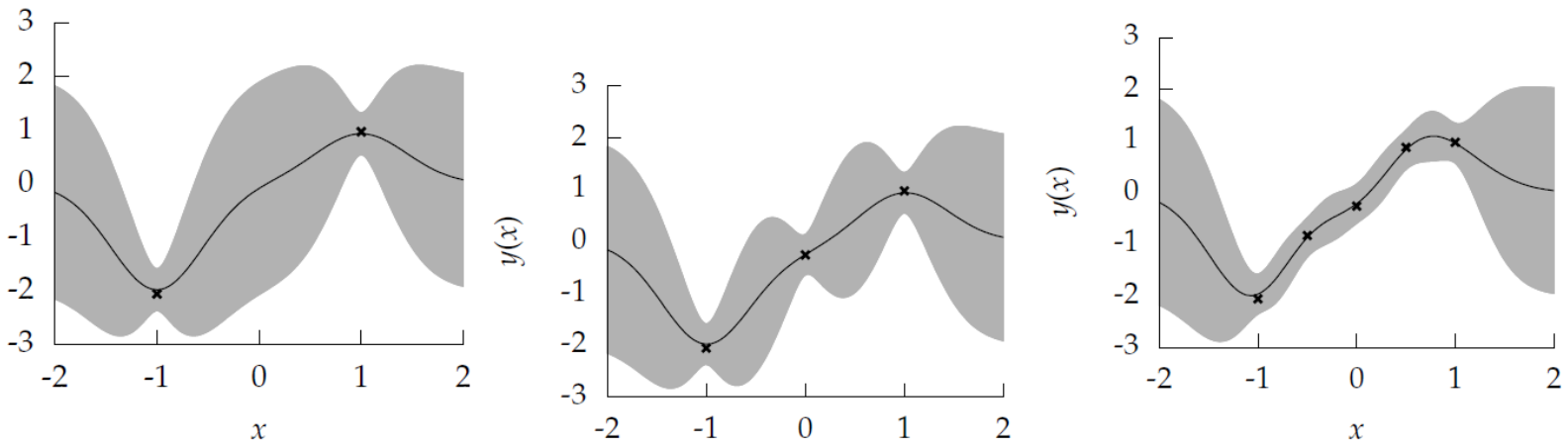
Thus we get

$$\mu_{\mathbf{X}_1|\mathbf{X}_2} = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$$

The important thing here is that we not only get a mean for each observed value, but we can also get the

The important thing here is that we not only get a mean for each observed value, but we can also get the variances from the diagonals of $\Sigma_{\mathbf{X}_1|\mathbf{X}_2}$. Each variance acts as a measure of the uncertainty in that prediction. It gauges how accurate the prediction is, and shows the degree to which an actual observation could depart from the predicted mean.

To visualize this uncertainty, a band can be drawn around the predicted mean that extends one or two standard deviations above and below the mean. An example is shown below for three sets of observations.



34: An example of a estimations on a line segment, with $N - d = 2$ (left), $N - d = 3$ (middle), and $N - d = 5$ (right) observations as black points. The predicted mean is shown as a black line, the grey band is two standard deviations away from the mean, using the computed standard deviation for the prediction.

You will notice that as more observations are added, the band shrinks in size. This shows that more data makes for a more precise prediction.

You will also notice that in the above figure, there is no hint of the discretization that you may expect from a finite number of prediction variables. Instead of choosing some large d and computing the full $d \times d$ covariance matrix, we instead set $d = 1$. This results in just a single prediction, but we are free to move the location of this prediction around. The above plot is made by scanning this single prediction along the x-axis, and plotting the mean and standard deviation as a function of the prediction location x.

Quantifying Uncertainty 1

1 point possible (graded)

If one could select the next point to take a measurement, is the mid-point between two existing points the best place to make this measurement? That is, the best place in terms of the maximal reduction of the shaded area.

- Yes
- No
- Not Necessarily

Submit You have used 0 of 1 attempt

Quantifying Uncertainty 2

1 point possible (graded)

Is the shaded gray area symmetric? That is, the same width above and below the black line?

- Yes
- No

Submit You have used 0 of 1 attempt

Quantifying Uncertainty 3

1 point possible (graded)

Is it possible that at some point in our estimation, when observing a new point on the line, the width of the shaded gray area increases for some values of the x-axis?

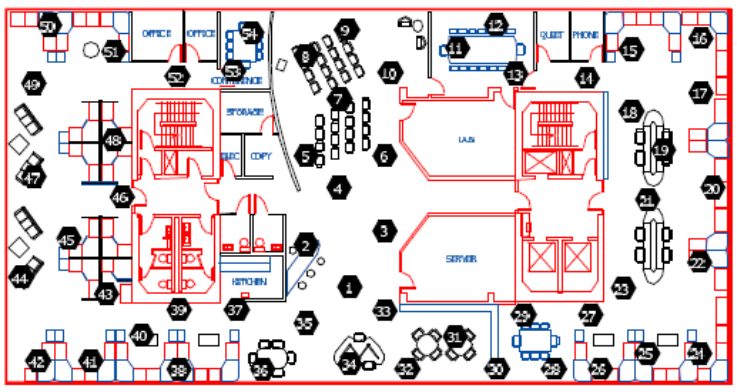
- Yes
- No

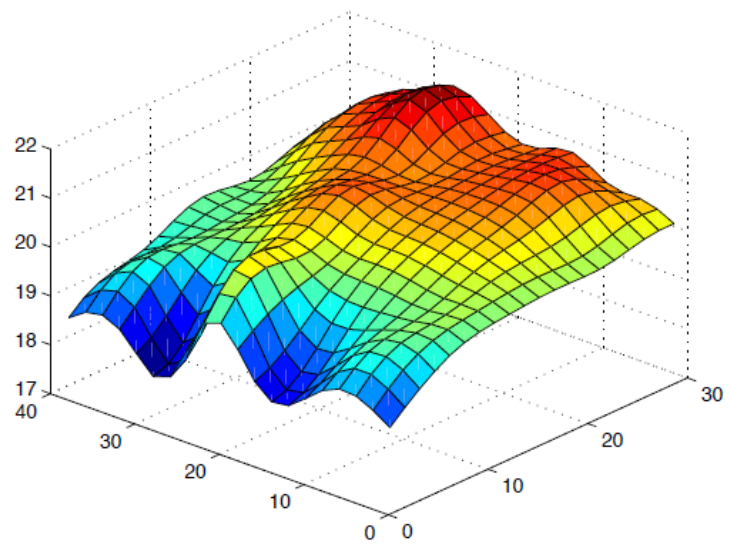
Submit

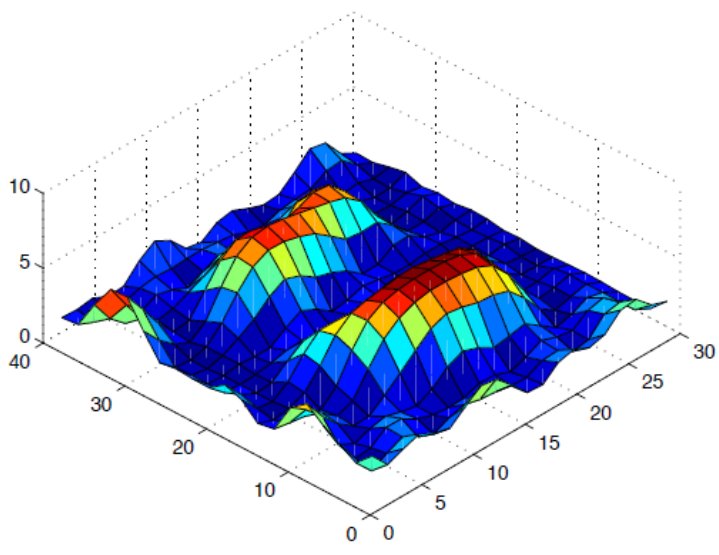
You have used 0 of 1 attempt

The below figure shows another example of the effects of the spatial correlations for temperature measurements. In the top image, a diagram of a classroom is shown, where the location of the temperature sensors is shown as black hexagons. In the two bottom images, the corresponding estimates for mean and variances are shown. This will be a two-dimensional example of what we have shown in the example above.

Note that in particular, the area where it is less dense in terms of the number of sensors is precisely the area where the highest variance of the estimates occurs. This is expected, as the covariance will decrease with distance, the areas which are effectively the farthest to a sensor will be the ones with higher uncertainty about their estimated values. Of course, those are not necessarily the precise points that are the farthest in the distance, as they need to be weighted by their corresponding variances.







(a) *Temperature prediction using GP* (b) *Variance of temperature prediction*

35: An example of a estimations on a 2D spatial dataset.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Spatial Prediction / 3. How to Model Covariance Matrices

Add a Post

Show all posts by recent activity

?	[Staff] Assumption of Normal distribution and Fig34 It seems that Fig 34 is not drawn with the assumption of Normal distribution of the random variables. The give-away is that the blac...	2
?	Figure 34 interpretation In figure 34, when taking additional measurements between -1 and 1, it seems that shaded area before -1 and after 1 did not change....	1
☑	[Staff] a little mistake: "column vector" instead of "row vector" at 17:42	3

‹ Previous Next ›