2. Recap

In the previous sections, we have built knowledge for estimating unobserved random variables based on the realization of other correlated variables. In particular, we have exploited correlations in space to make predictions about unobserved events. For example, in a simplistic model, we have assumed that the temperatures in two cities are correlated based on their proximity. Of course, many other phenomena also play a role, like altitude, geography, etc., but we make such simplification for clarity of exposition. Then, if one has sufficient statistics about both cities' temperature, once we observe the temperature in one of them, we can estimate the temperature in the other. Such simplifications could also be reasonable for other environmental variables like precipitation volume, underground oil reserves, etc.

Our generic problem assumes that we have $N-d$ random variables, denoted as $\mathbf{X}_2 \in \mathbb{R}^{N-d}$, with observed realizations (also referred to as observations or measurements) denoted by its lower case $\mathbf{x}_2 \in \mathbb{R}^{N-d}$. Furthermore, let us assume that each of the observations corresponds to some geographical location, and lets call such position $\mathbf{Z}_2 \in \mathbb{R}^{M\times(N-d)}$ assuming the spaces are in $\mathbb{R}^M$ (for example, we might have that $M=2$ and $\mathbf{Z}_2$ is a matrix of geospatial coordinates.)

Now, we have a new set of locations $\mathbf{Z}_1 \in \mathbb{R}^{M\times d}$, and we would like to estimate its temperatures $\mathbf{X}_1 \in \mathbb{R}^d$. Our main assumption is that the random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ are jointly Gaussian, therefore,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \in \mathbb{R}^d \\ \mathbf{X}_2 & \in \mathbb{R}^{N-d} \end{bmatrix}$$

where

$$\mu_{\mathbf{X}} = \begin{bmatrix} \mu_1 & \in \mathbb{R}^d \\ \mu_2 & \in \mathbb{R}^{N-d} \end{bmatrix} \text{ and } \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} \in \mathbb{R}^{d\times d} & \mathbf{\Sigma}_{12} \in \mathbb{R}^{d\times(N-d)} \\ \mathbf{\Sigma}_{21} \in \mathbb{R}^{(N-d)\times d} & \mathbf{\Sigma}_{22} \in \mathbb{R}^{N\times N} \end{bmatrix}$$

Note that in order to estimate the values of the unobserved random variables $\mathbf{X}_1$, we need to know $\mu_{\mathbf{X}}$ and $\mathbf{\Sigma}$.

Finally, let us recall the meaning of the variables the table below.

A summary of the symbols used, their dimension, their meaning, and availability

Symbol	Dimension	Meaning	Availability
$\mathbf{X}_1$	$\mathbb{R}^d$	Unobserved random variables, e.g., temperatures at cities for which we do not have measurements but we would like to estimate.	We know $\mathbb{E}[\mathbf{X}_1]$ and $var(\mathbf{X}_1)$
$\mathbf{Z}_1$	$\mathbb{R}^{M\times d}$	Locations of the cities corresponding to $\mathbf{X}_1$ measurements.	Available
$\mathbf{X}_2$	$\mathbb{R}^{N-d}$	Observed random variables, e.g., temperatures at cities for which we have measurements.	We know $\mathbb{E}(\mathbf{X}_2)$ and $var(\mathbf{X}_2)$
$\mathbf{Z}_2$	$\mathbb{R}^{M\times(N-d)}$	Locations of the cities corresponding to $\mathbf{X}_2$ measurements.	Available
μ_1	$\mathbb{R}^d$	Mean of the unobserved random variables $\mathbf{X}_1$	We know $\mathbb{E}(\mathbf{X}_1) = \mu_1$

μ_2	$\mathbb{R}^(N-d)$	Mean of the observed random variables $\mathbf{X}_2$	We know $\mathbb{E}(\mathbf{X}_2) = \mu_2$
$\mathbf{x}_1$	$\mathbb{R}^d$	Realizations of the random variable $\mathbf{X}_1$	Not available
$\mathbf{x}_2$	$\mathbb{R}^{N-d}$	Realizations of the random variable $\mathbf{X}_2$	Available
Σ_{11}	$\mathbb{R}^{d\times d}$	Variance of the random variables $\mathbf{X}_1$	Available $var(\mathbf{X}_1)$.
Σ_{12}	$\mathbb{R}^{d\times(N-d)}$	Covariance of the random variables $\mathbf{X}_1$ with $\mathbf{X}_2$	Available $cov(\mathbf{X}_1, \mathbf{X}_2)$.
Σ_{21}	$\mathbb{R}^{(N-d)\times d}$	Covariance of the random variables $\mathbf{X}_2$ with $\mathbf{X}_1$	Available $cov(\mathbf{X}_2, \mathbf{X}_1)$.
Σ_{22}	$\mathbb{R}^{(N-d)\times(N-d)}$	Variance of the random variables $\mathbf{X}_2$	Available $var(\mathbf{X}_2)$.
$\mu_{\mathbf{X}_1\vert\mathbf{X}_2}$	$\mathbb{R}^d$	New estimate of the mean of $\mathbf{X}_1$ given the observations $\mathbf{x}_2$ of the random variable $\mathbf{X}_2$	We compute this.
$\Sigma_{\mathbf{X}_1\vert\mathbf{X}_2}$	$\mathbb{R}^{d\times d}$	New estimate of the variance of $\mathbf{X}_1$ given the observations $\mathbf{x}_2$ of the random variable $\mathbf{X}_2$	We compute this.

In the next section, we will explore how to model the covariance matrix $\mathbf{\Sigma}$.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Spatial Prediction / 2. Recap

< Previous

Next >

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog