Home Assignment - 1

Santhosh Nadig

November 17, 2016

1 Introduction

2 Theory

Let $\mathbf{Y}_k$ represent the n known observations of spatial data and $\mathbf{Y}_u$ represent the m data points that needs to be estimated. For ordinary Kriging ($\mu = \mathbf{1}\beta$), the optimal predictions are given by

$$\begin{aligned}\hat{\mathbf{Y}}_u &= \mathbf{1}_u\hat{\beta} + \Sigma_{uk}\Sigma_{kk}^{-1}\left(\mathbf{Y}_k - \mathbf{1}_k\hat{\beta}\right)\\ \hat{\beta} &= \left(\mathbf{1}_k^T\Sigma_{kk}^{-1}\mathbf{1}_k\right)^{-1}\mathbf{1}_k^T\Sigma_{kk}^{-1}\mathbf{Y}_k,\end{aligned}$$

where

$$\begin{bmatrix}\mathbf{Y}_k\\ \mathbf{Y}_u\end{bmatrix} \in \mathbf{N}\left(\begin{bmatrix}\mathbf{1}_k\beta\\ \mathbf{1}_u\beta\end{bmatrix}, \begin{bmatrix}\Sigma_{kk} & \Sigma_{ku}\\ \Sigma_{uk} & \Sigma_{uu}\end{bmatrix}\right).$$

Firstly, we show that the predictions are linear in the observations, i.e., $\hat{\mathbf{Y}}_u = \boldsymbol{\lambda}^T\mathbf{Y}_k$ for some $\boldsymbol{\lambda}$. We may re-write $\hat{\beta}$ as

$$\begin{aligned}\hat{\beta} &= \boldsymbol{\gamma}^T\mathbf{Y}_k\\ \boldsymbol{\gamma} &= \left(\left(\mathbf{1}_k^T\Sigma_{kk}^{-1}\mathbf{1}_k\right)^{-1}\mathbf{1}_k^T\Sigma_{kk}^{-1}\right)^T.\end{aligned}$$

We note that the vector $\boldsymbol{\gamma}^T$ is of dimensions $1 \times n$, and therefore, $\hat{\beta}$ is a scalar. Substituting the above in the expression for $\hat{\mathbf{Y}}_u$, we obtain

$$\begin{aligned}\hat{\mathbf{Y}}_u &= \mathbf{1}_u\boldsymbol{\gamma}^T\mathbf{Y}_k + \Sigma_{uk}\Sigma_{kk}^{-1}\mathbf{Y}_k - \Sigma_{uk}\Sigma_{kk}^{-1}\mathbf{1}_k\boldsymbol{\gamma}^T\mathbf{Y}_k\\ &= \left(\mathbf{1}_u\boldsymbol{\gamma}^T + \Sigma_{uk}\Sigma_{kk}^{-1} - \Sigma_{uk}\Sigma_{kk}^{-1}\mathbf{1}_k\boldsymbol{\gamma}^T\right)\mathbf{Y}_k\\ &= \boldsymbol{\lambda}^T\mathbf{Y}_k,\end{aligned}$$

where $\boldsymbol{\lambda} = \left(\mathbf{1}_u\boldsymbol{\gamma}^T + \Sigma_{uk}\Sigma_{kk}^{-1} - \Sigma_{uk}\Sigma_{kk}^{-1}\mathbf{1}_k\boldsymbol{\gamma}^T\right)^T$.

Secondly, we show that the predictions are unbiased, i.e., $\mathbf{E}(\hat{\mathbf{Y}}_u) = \mathbf{E}(\mathbf{Y}_u)$. The expected value of $\hat{\beta}$ is given by

$$\begin{aligned}\mathbf{E}\left(\hat{\beta}\right) &= \left(\mathbf{1}_k^T\Sigma_{kk}^{-1}\mathbf{1}_k\right)^{-1}\mathbf{1}_k^T\Sigma_{kk}^{-1}\ \mathbf{E}(\mathbf{Y}_k)\\ &= \left(\mathbf{1}_k^T\Sigma_{kk}^{-1}\mathbf{1}_k\right)^{-1}\mathbf{1}_k^T\Sigma_{kk}^{-1}\ \mathbf{1}_k\beta\\ &= \beta.\end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbf{E}(\hat{\mathbf{Y}}_u) &= \mathbf{1}_u \mathbf{E}\left(\hat{\beta}\right) + \Sigma_{uk}\Sigma_{kk}^{-1}\left(\mathbf{E}(\mathbf{Y}_k) - \mathbf{1}_k \mathbf{E}\left(\hat{\beta}\right)\right) \\
&= \mathbf{1}_u \beta + \Sigma_{uk}\Sigma_{kk}^{-1}\left(\mathbf{1}_k\beta - \mathbf{1}_k\beta\right) \\
&= \mathbf{1}_u \beta.
\end{aligned}
$$

Finally, consider a second unbiased predictor of $\mathbf{Y}_u$, given by $\tilde{\mathbf{Y}}_u = (\boldsymbol{\lambda} + \boldsymbol{\nu})^T\mathbf{Y}_k$. From the unbiasedness criterion, we have that $\mathbf{E}(\tilde{\mathbf{Y}}_u) = \mathbf{E}(\hat{\mathbf{Y}}_u)$. Therefore, $(\boldsymbol{\lambda} + \boldsymbol{\nu})^T\mathbf{E}(\mathbf{Y}_k) = \boldsymbol{\lambda}^T\mathbf{E}(\mathbf{Y}_k)$, implying $\boldsymbol{\nu}^T\mathbf{1}_k = 0$. Furthermore, consider the variance of the second predictor

$$
\begin{aligned}
\mathbf{V}(\tilde{\mathbf{Y}}_u - \mathbf{Y}_u) &= \mathbf{V}(\boldsymbol{\nu}^T\mathbf{Y}_k + (\boldsymbol{\lambda}^T\mathbf{Y}_k - \mathbf{Y}_u)) \\
&= \mathbf{V}(\boldsymbol{\nu}^T\mathbf{Y}_k) + \mathbf{V}(\hat{\mathbf{Y}}_u - \mathbf{Y}_u) + 2\mathbf{C}(\boldsymbol{\nu}^T\mathbf{Y}_k, \hat{\mathbf{Y}}_u - \mathbf{Y}_u) \\
&\geq \mathbf{V}(\hat{\mathbf{Y}}_u - \mathbf{Y}_u) + 2\mathbf{C}(\boldsymbol{\nu}^T\mathbf{Y}_k, \hat{\mathbf{Y}}_u - \mathbf{Y}_u) \\
&= \mathbf{V}(\hat{\mathbf{Y}}_u - \mathbf{Y}_u) + 2\boldsymbol{\nu}^T\mathbf{C}(\mathbf{Y}_k, \hat{\mathbf{Y}}_u) - 2\boldsymbol{\nu}^T\mathbf{C}(\mathbf{Y}_k, \mathbf{Y}_u) \\
&= \mathbf{V}(\hat{\mathbf{Y}}_u - \mathbf{Y}_u) + 2\boldsymbol{\nu}^T\left(\Sigma_{kk}\boldsymbol{\lambda} - \Sigma_{uk}\right)
\end{aligned}
$$

Thus, if we can choose $\boldsymbol{\lambda}$ such that $\boldsymbol{\nu}^T\left(\Sigma_{kk}\boldsymbol{\lambda} - \Sigma_{uk}\right) = 0$, then, we have that $\mathbf{V}(\tilde{\mathbf{Y}}_u - \mathbf{Y}_u) \geq \mathbf{V}(\hat{\mathbf{Y}}_u - \mathbf{Y}_u)$ for any unbiased predictor $\tilde{\mathbf{Y}}_u$.

3 Swedish Temparature Reconstruction

The data consists of average temperature during June 2005 measured at 250 stations across Sweden. In addition, the latitude, longitude, elevation and distance to both the Swedish coast and to any coastline are provided too. Figure 1 shows the plot of average temperature, elevation and distance to any coast. The average temperatures are higher in the South (lower latitudes) and towards the coast, as expected.

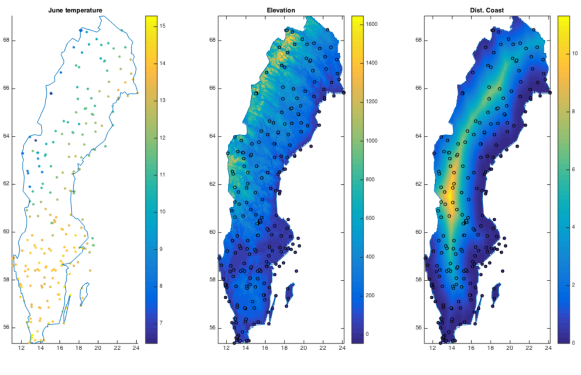


Figure 1: Raw Data Plot...

3.1 Least-Squares

The mean of the spatial data is modelled as a linear function of the covariates. Figure 2 shows a plot of the mean temperature (observations) as a function of the available covariates from the dataset `SweObs` containing the 250 measurements. It was found that longitude does not relate linearly with the average

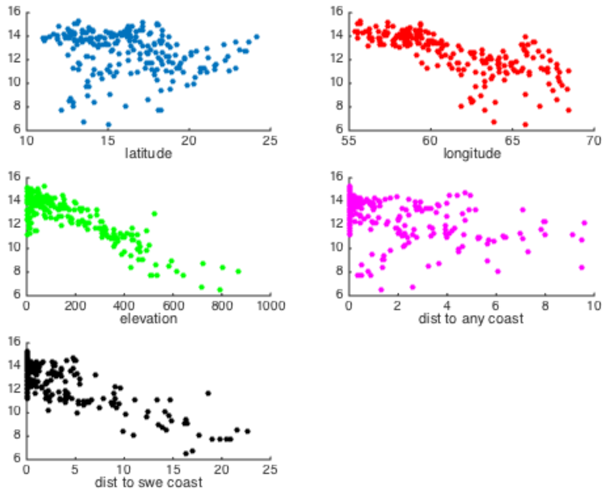


Figure 2: Observation vs covariates

temperatures measured and hence is left out of modelling. In order to select the appropriate covariates, we formed five models (A,B,C,D,E) containing different set of covariates as tabulated in Table 1.

Model	Covariates
* A	Latitude, elevation, distance to any coast, distance to Swedish coast
B	Latitude, elevation, distance to Swedish coast
C	Latitude, elevation, distance to any coast
D	Latitude, distance to any coast, distance to Swedish coast
E	Latitude, elevation

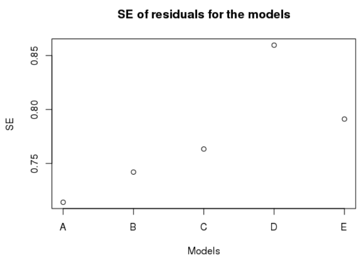
Table 1: Models and the respective covariates, with average temperature as the independent varible.

The standard errors of the residuals and the Akaike Information Criterion (AIC) seen in figure 3 shows that model A performs best.

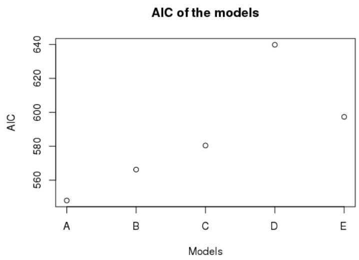
Thus, using model A, we may express the observed temperature $\boldsymbol{\mu}$ as a linear function of latitude ($\mathbf{x}_1$), elevation ($\mathbf{x}_2$), distance to any coast ($\mathbf{x}_3$) and the distance to Swedish coast ($\mathbf{x}_4$).

$$\boldsymbol{\mu} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + +\beta_4\mathbf{x}_4 + \boldsymbol{\epsilon} \tag{1}$$

where β_0 is the intercept, $\beta_1 \ldots \beta_4$ are the parameters and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$. The estimated parameters (and their standard deviations) are tabulated below.



(a) Std. Err of the residuals for each of the models.



(b) AIC for each of the models.

Figure 3: Performance evaluation of the five models (A,B,C,D,E). Model A has the least std. error of the residuals and the best AIC metric.

Parameter	Estimate	SE
$\hat{\beta}_0$	24.023	0.9
$\hat{\beta}_1$	-0.169	0.015
$\hat{\beta}_2$	-0.005	0.0004
$\hat{\beta}_3$	0.115	0.025
$\hat{\beta}_4$	-0.09	0.025

Table 2: Parameter Estimates and corresponding standard errors.

The variance of the residuals ${\sigma_\epsilon}^2 = 0.5098$.

3.2 Universal Kriging

4 Conclusions