Home Assignment - 1

Santhosh Nadig

November 19, 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 dataset (`SweObs`) consists of average temperature during June 2005 measured at 250 stations across Sweden. The corresponding 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.

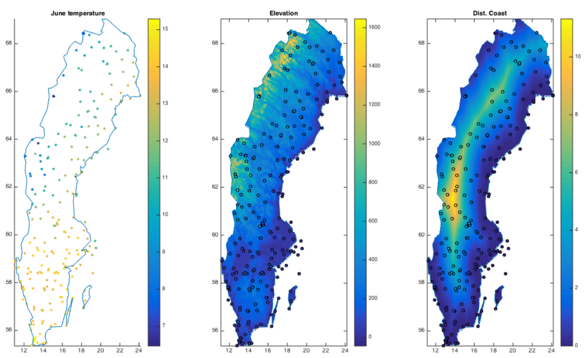


Figure 1: Plots showing (R-L) average temperature in June 2005, elevation and distance to any coastline.

3.1 Ordinary Least Squares

The average temperature (observations), $\mathbf{Y}$ is modelled as a linear function of the covariates.

$$\mathbf{Y} \triangleq \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{X}$ is the covariate matrix, $\boldsymbol{\beta}$ is the parameter vector and $\boldsymbol{\epsilon} \in N(0, {\sigma_\epsilon}^2\mathbb{I})$.

Figure 2 shows a plot of the observations as a function of the available covariates from `SweObs`. It was found that longitude does not relate linearly

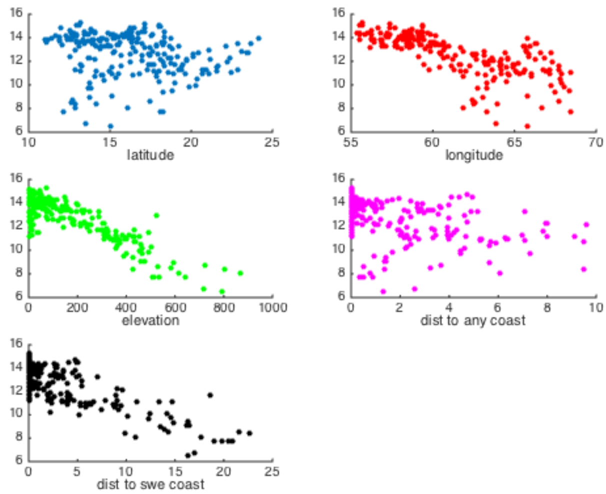


Figure 2: Observation vs covariates

with the average 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.

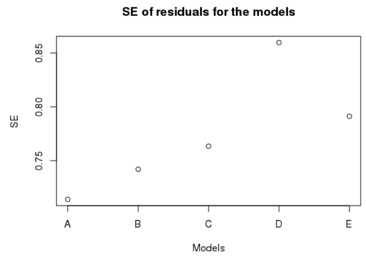
Model	**Covariates**
* A	Latitude, elevation, distance to any coast, distance to Swedish coast
B	Latitude, elevation, distance to Swedish coast
C	Latitude, elevation, distance to any coast
D	Latitude, distance to any coast, distance to Swedish coast
E	Latitude, elevation

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

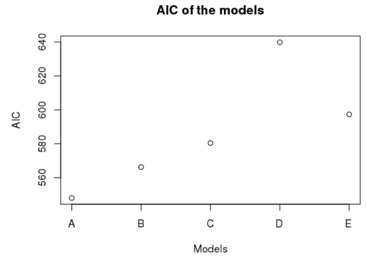
The standard errors of the residuals and the Akaike Information Criterion (AIC) seen in figure 3 shows that model A performs best.

Thus, using model A, we may express the observations $\mathbf{Y}$ 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 ($\mathbf{x}_4$).

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



(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. Err of the residuals and the best AIC metric.

where $\beta_0 \ldots \beta_4$ are the parameters. For validation, 25 randomly picked observations are set aside and we use $N = 225$ data points for the model parameter estimation. The estimated parameters and variance are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
$$\mathbf{V}(\hat{\boldsymbol{\beta}}|{\sigma_\epsilon}^2) = {\sigma_\epsilon}^2(\mathbf{X}^T\mathbf{X})^{-1}$$

The estimated parameters (and their standard errors) are tabulated below.

Parameter	**Estimate**	**SE**
$\hat{\beta}_0$	23.972	0.95
$\hat{\beta}_1$	-0.169	0.016
$\hat{\beta}_2$	-0.005	0.0005
$\hat{\beta}_3$	0.106	0.025
$\hat{\beta}_4$	-0.089	0.015

Table 2: Parameter Estimates and corresponding standard errors (OLS).

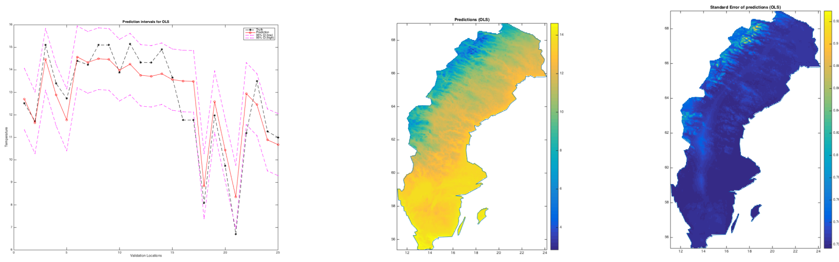
The variance of the residuals ${\sigma_\epsilon}^2 = 0.48$. The mean temperatures at the validation points are given by

$$\mathbf{E}(\hat{\mathbf{Y}}_v) = \mathbf{X}_v\hat{\boldsymbol{\beta}} \tag{2}$$

where $\hat{\mathbf{Y}}_v$ denotes the average temperature estimates at the validation points, $\mathbf{X}_v$ the regressor matrix formed using the latitude, elevation, distance to any coast and distance to Swedish coast of the 25 validation points and $\hat{\boldsymbol{\beta}}$ the estimated model parameters. The variance and the 95% confidence interval are given by

$$\mathbf{V}(\hat{\mathbf{Y}}_v) = {\sigma_\epsilon}^2(\mathbf{X}_v(\mathbf{X}_v^T\mathbf{X}_v)^{-1}\mathbf{X}_v^T + 1)$$
$$\text{CI, } 95\% = \mathbf{E}(\hat{\mathbf{Y}}_v) \pm 1.96 \text{ SE}(\hat{\mathbf{Y}}_v)$$

where $\text{SE}(\hat{\mathbf{Y}}_v) = \sqrt{\mathbf{V}(\hat{\mathbf{Y}}_v)}$. Figure 4 shows a plot of prediction and confidence intervals for the validation points and also the predictions for the `SweGrid` data along with the corresponding standard error.



(a) (b) (c)

Figure 4: (a): Estimate and confidence intervals (95%) for the 25 validation points. (b),(c): Predictions and SE of predictions for `SweGrid` data.

3.2 Universal Kriging

Firstly, we proceed to evaluate the spatial dependence in the residuals from the OLS model. A non-parametric approach (binned covariance function estimation [slide 13 lecture 03]) with 32 bins is used for this purpose. Figure shows the estimated covariance function of the residuals (as a function of the bins), where the bin $\mathbf{H}_k = \{(i,j) : i \neq j, kh \leq \|\mathbf{s}_j - \mathbf{s}_i\| < (k+1)h\}$. The confidence intervals are generated by the boxplot where the first and the third quantiles (25, 75) are shown. Thus, we see that there are five points outside the quantiles implying a significant covariance structure.

The covariance function is modelled as a Matern covariance function whose parameters are estimated using the least squares method. That is, given the residuals z_i

4 Conclusions