Project : Sequential Uncertainty Reduction and Value of Information in Gaussian Process Models

Reconsider the situation where a farmer must decide whether to harvest a forest with uncertain profits, or just leave it for conservation and receive 0 profits for sure. The forest is split in 625 units represented on a regular 25×25 grid cells: $(1,1),(1,2),\ldots,(25,25)$. We assume a setting with low decision flexibility, so the farmer has just two choices $a = 0$: leave all cells for conservation, $a = 1$: harvest all cells.

The uncertain profits are models by a Gaussian process (GP) represented on the grid. The prior mean at a cell $\boldsymbol{s} = (s_1, s_2)$ in the grid is $E(x(\boldsymbol{s})) = 0$, while the variance is $\text{Var}(x(\boldsymbol{s})) = 1^2$ and correlation $\text{Corr}(x(\boldsymbol{s}), x(\boldsymbol{s}')) = e^{-0.15h}$ for Euclidean distance $h = \sqrt{||\boldsymbol{s} - \boldsymbol{s}'||^2}$. Using vector form, $\boldsymbol{x} = (x_1, \ldots, x_{625})$ represents the profits at all locations, and this vector has multivariate Gaussian prior distribution with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ having entries defined by the functions above.

Surveying can be done along North-South lines in the grid. If a line is selected for data gathering, measurements are made at each of the 25 cells in this line, with model $y(\boldsymbol{s}) = x(\boldsymbol{s}) + N(0, 5^2)$. When line $j \in \{1, \ldots, 25\}$ is chosen, the measurement model can be written

$$\boldsymbol{y}_j = \boldsymbol{F}_j \boldsymbol{x} + N(\mathbf{0}, \boldsymbol{T}),$$

where $\boldsymbol{F}_j$ is a 25×625 matrix which picks the entries of the vector $\boldsymbol{x}$ corresponding to the jth line, and $\boldsymbol{T}$ is a 25×25 diagonal matrix with entries 5^2.

If data is collected, the updated (posterior) distribution is Gaussian with new mean and covariance

$$\boldsymbol{\mu}_j = \boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{F}_j^t(\boldsymbol{F}_j\boldsymbol{\Sigma}\boldsymbol{F}_j^t + \boldsymbol{T})^{-1}(\boldsymbol{y}_j - \boldsymbol{F}_j\boldsymbol{\mu})$$

$$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{F}_j^t(\boldsymbol{F}_j\boldsymbol{\Sigma}\boldsymbol{F}_j^t + \boldsymbol{T})^{-1}\boldsymbol{F}_j\boldsymbol{\Sigma}$$

We will next focus on sequential surveying of lines, where the posterior at one stage forms the prior for the next stage.

a)

Select the line j that maximises the reduction in the mean variance of the grid: i.e. the one that gives the smallest $(1/625)\sum_{k=1}^{625} \Sigma_j(k,k)$.

Conduct sequential selection of lines, where the covariances is updated for each stage. Which 10 lines are selected first?

b)

For the decision situation with harvest ($a = 1$) or not harvest ($a = 0$), one is interested in worthwhile information gathering for making improved decisions.

The prior value is given by $PV = \max\{0, \sum_{k=1}^{625} \mu(k)\}$.

For the first stage the value of information (VOI) is:

$$\text{VOI}_j = [\mu_w \Phi(\mu_w / r_{w,j}) + r_{w,j} \phi(\mu_w / r_{w,j})] - PV,$$

using the Gaussian modeling assumptions. Here, $\mu_w = \sum_{k=1}^{625} \mu(k)$ while $r_{w,j}^2 = \sum_k \sum_l R_j(k,l)$ and $\boldsymbol{R}_j = \boldsymbol{\Sigma} \boldsymbol{F}_j^t (\boldsymbol{F}_j \boldsymbol{\Sigma} \boldsymbol{F}_j^t + \boldsymbol{T})^{-1} \boldsymbol{F}_j \boldsymbol{\Sigma}$. Compute and compare VOI_j for all 25 North-South lines.

At later stages one will continue to gather data sequentially (at the next best line, given the current information), if the value is larger than the price of gathering one more line. Otherwise, the decision maker will stop the data gathering.

Play the game by generating data at the selected lines and gather data until the value of more sequential information is less than price $P = 0.5$. Repeat this for 100 replicate trials. Make a histogram of the number of lines selected and plot which lines are selected at the different replicate trials.