Lecture 11
More about Kriging

Dennis Sun
Stats 253

July 28, 2014

Outline of Lecture

1. Kriging and Isotropy
2. Kriging for Spatio-Temporal Data
3. Other Interpretations of Kriging
4. Wrapping Up

Where are we?

1. Kriging and Isotropy
2. Kriging for Spatio-Temporal Data
3. Other Interpretations of Kriging
4. Wrapping Up

Isotropy

Covariance Function: $\Sigma(\boldsymbol{s}, \boldsymbol{s}')$

- Original problem: covariance between $y(\boldsymbol{s})$ and $y(\boldsymbol{s}')$ can be "anything".
- Stationarity: covariance only depends on difference between points:

$$\Sigma(\boldsymbol{s}, \boldsymbol{s}') = C(\boldsymbol{s} - \boldsymbol{s}')$$

- Isotropy: covariance only depends on distance between points:

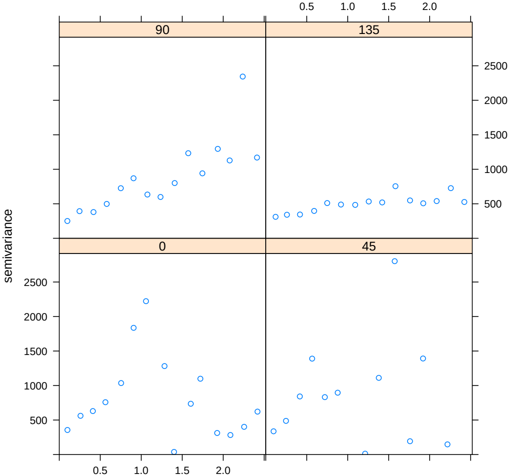
$$\Sigma(\boldsymbol{s}, \boldsymbol{s}') = C(||\boldsymbol{s} - \boldsymbol{s}'||)$$

- How did isotropy help us when kriging?

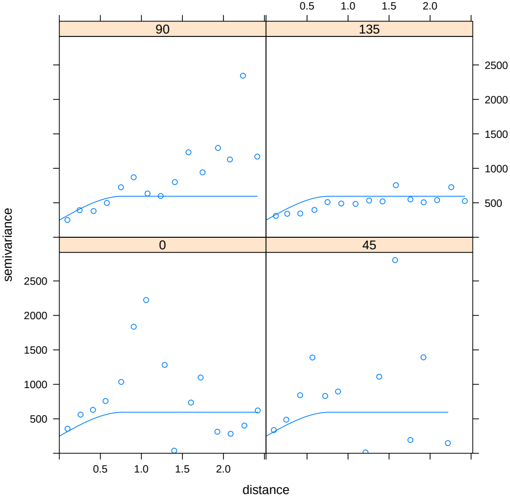
Examining for (An)isotropy

Calculate separate variograms for different directions.

```
sample.vgm <- variogram(snow_wet ~ elevation + latitude, data=data,
                        alpha=c(0,45,90,135))
```

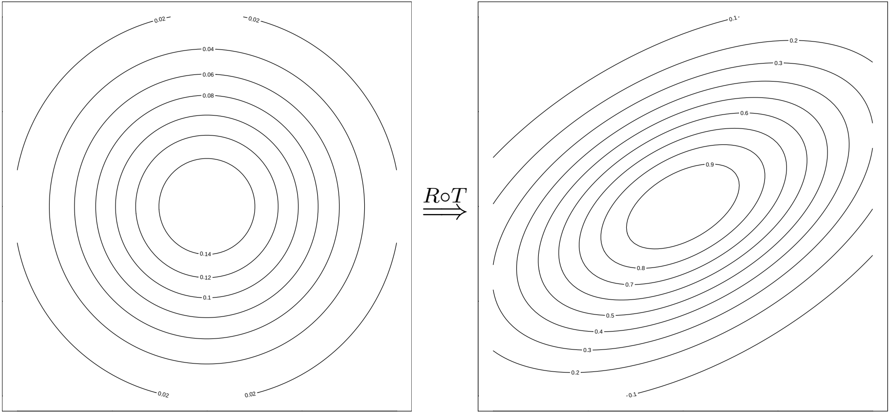


Isotropy seems reasonable on the snowpack data



What if isotropy is not reasonable?

Can construct anisotropic covariance as a shear of isotropic one.



$$C_{an}(\boldsymbol{s} - \boldsymbol{s}') = C_{is}(T^{-1}R^{-1}(\boldsymbol{s} - \boldsymbol{s}'))$$

Where are we?

1. Kriging and Isotropy
2. Kriging for Spatio-Temporal Data
3. Other Interpretations of Kriging
4. Wrapping Up

A Warning

- This section is going to be extremely boring.
- Kriging in space-time is pretty much exactly the same as kriging in space, with an extra dimension.
- Is $2 + 1 = 3$? (For spatio-temporal kriging, more or less.)

Definitions

- We now observe $y(\boldsymbol{s}, t)$.
- The covariance function is now

$$\Sigma((\boldsymbol{s}, t), (\boldsymbol{s}', t')) = \mathrm{E}\left[y(\boldsymbol{s}, t) y(\boldsymbol{s}', t')\right] - \mu(\boldsymbol{s}, t)\mu(\boldsymbol{s}', t').$$

- Stationarity means

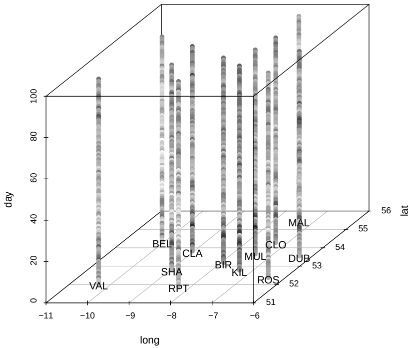
$$\Sigma((\boldsymbol{s}, t), (\boldsymbol{s}', t')) = C(\boldsymbol{s} - \boldsymbol{s}', t - t').$$

- How should isotropy be defined?
- Should probably keep isotropy in space separate from isotropy in time.

$$\Sigma((\boldsymbol{s}, t), (\boldsymbol{s}', t')) = C(||\boldsymbol{s} - \boldsymbol{s}'||, |t - t'|).$$

Separability

Assume stationarity.



Irish wind data

- Easiest way is to assume a **separable** model:

$$C(\boldsymbol{r}, h) = C_S(\boldsymbol{r})C_T(h)$$

- Suggests we can fit covariance function to spatial and temporal components separately:

$$\hat{C}(\boldsymbol{r}, h) = \hat{C}_S(\boldsymbol{r})\hat{C}_T(h)$$

- Fully symmetric:

$$C(\boldsymbol{r}, h) = C(\boldsymbol{r}, -h)$$

Is this reasonable? (Consider a wind that blows west.)

Non-Separable Covariances

- Gneiting (2002) proposes class of functions:

$$C(\boldsymbol{r}, h) = \frac{\sigma^2}{(|h|^{2\gamma} + 1)^\tau} \exp\left[\frac{-c||\boldsymbol{r}||^{2\gamma}}{(|h|^{2\gamma} + 1)^{\beta\gamma}}\right].$$

  Strength of spatio-temporal correlation governed by β.

- Another approach is to take a weighted combination of covariance functions (De Iaco et al, 2001):

$$C(\boldsymbol{r}, h) = w_0 C_S(\boldsymbol{r}) C_T(h) + w_1 \tilde{C}_S(\boldsymbol{r}) + w_2 \tilde{C}_T(h).$$

Summary of Spatio-Temporal Kriging

- For data in continuous space and time, it pretty much the only game in town.
- How often do you have continuous observations over both space and time anyway? Typically, you observe snapshots of spatial data at discrete times.
- If time is discrete, then we can model the temporal component using, e.g., AR process.
- **Next class:** how to do this. (Hint: it involves the Kalman filter.)

Where are we?

1 Kriging and Isotropy

2 Kriging for Spatio-Temporal Data

3 Other Interpretations of Kriging

4 Wrapping Up

Kriging as Function Estimation

- Recall the kriging estimator

$$\hat{y}_0 = \Sigma_{y_0 \boldsymbol{y}} \Sigma_{\boldsymbol{y}\boldsymbol{y}}^{-1} \boldsymbol{y}.$$

- We can think of $\hat{y}_0$ as a function

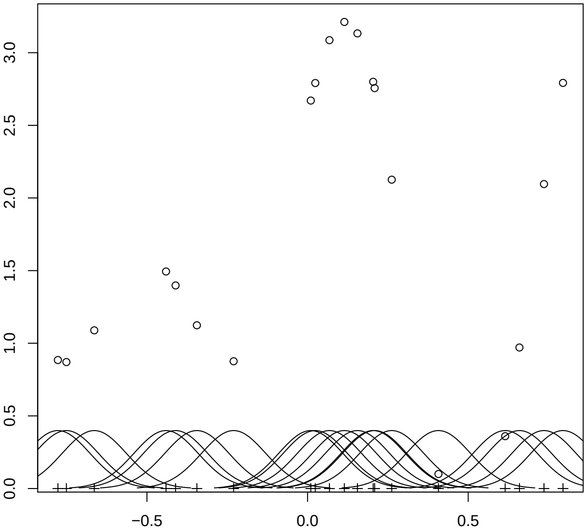
$$\hat{y}(\boldsymbol{s}_0) = \Sigma(\boldsymbol{s}_0, S)\Sigma(S, S)^{-1}\boldsymbol{y}$$

  where $S = \begin{bmatrix} \boldsymbol{s}_1 & \cdots & \boldsymbol{s}_n \end{bmatrix}$.
- Another way to write this is

$$\hat{y}(\boldsymbol{s}_0) = \sum_{i=1}^{n} \hat{c}_i \Sigma(\boldsymbol{s}_0, \boldsymbol{s}_i)$$

  where $\hat{\boldsymbol{c}} = \Sigma(S, S)^{-1}\boldsymbol{y}$.
- We call $\Sigma(\cdot, \boldsymbol{s}_i)$ a kernel function.

Kriging as Function Estimation

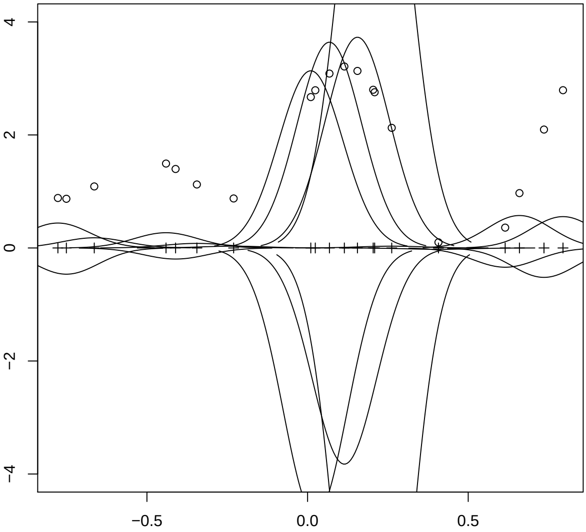


Kriging and Interpolation

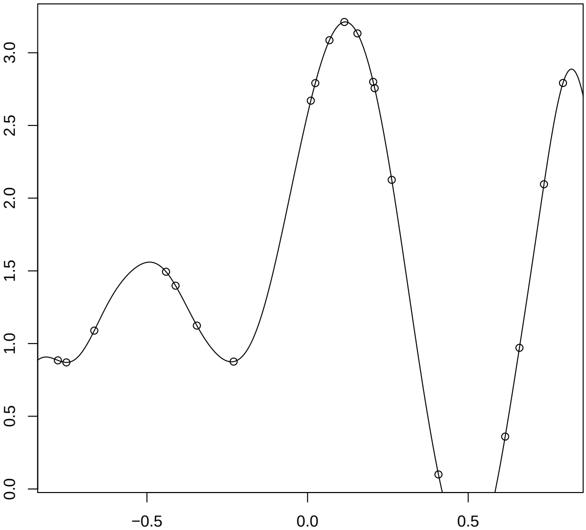
$$\hat{y}(\cdot) = \sum_{i=1}^{n} \hat{c}_i \Sigma(\cdot, \boldsymbol{s}_i) \qquad \hat{f}(\cdot) = \sum_{i=1}^{n} \hat{c}_i K(\cdot, \boldsymbol{s}_i)$$

- We have n unknowns c_i, $i = 1, ..., n$.
- We also have n constraints. What are they?
- Recall that $\hat{y}(\boldsymbol{s}_i) = \mathrm{E}(y(\boldsymbol{s}_i)|\boldsymbol{y}) = y(\boldsymbol{s}_i)$.
- So $\hat{f}$ has to pass through the n points.

Kriging and Interpolation



Kriging and Interpolation



One Application: Integration

Let's look at an application of this kind of interpolation:

- Suppose we want to integrate a function f, but we don't know f. We only observe it at a few points: $f(\boldsymbol{x}_i)$ for $i = 1, ..., n$.
- **Idea:** Interpolate points, then integrate interpolated function $\hat{f}$.
- What happens? Suppose $\int K(\boldsymbol{x}, \boldsymbol{x}')\, d\boldsymbol{x} = 1$ for any $\boldsymbol{x}'$.

$$\int \hat{f}(\boldsymbol{x})\, d\boldsymbol{x} = \int \sum_{i=1}^{n} \hat{c}_i K(\boldsymbol{x}, \boldsymbol{x}_i)\, d\boldsymbol{x} = \sum_{i=1}^{n} \hat{c}_i$$

- So to integrate, we just need to compute weights $\hat{c}_i$ and sum them. No numerical integration involved!

Adding Observation Noise

- Suppose we actually observe $z(\boldsymbol{s}_i) = y(\boldsymbol{s}_i) + \delta_i$, where $\delta_i \sim N(0, \tau^2)$.
- This changes the kriging estimator slightly

$$\hat{y}_0 = \Sigma_{y_0 \boldsymbol{y}}(\Sigma_{\boldsymbol{yy}} + \tau^2 I)^{-1}\boldsymbol{y}$$
$$\hat{y}(\cdot) = \sum_{i=1}^{n} \hat{c}_i \Sigma(\cdot, \boldsymbol{s}_i)$$

where $\hat{c}_i = (\Sigma(S,S) + \tau^2 I)^{-1}\boldsymbol{y}$.

- Now our data is generated from the Gaussian process model:

$$y(\cdot) \sim GP(0, \Sigma(\cdot,\cdot))$$
$$z(\boldsymbol{s}_i) = y(\boldsymbol{s}_i) + \delta_i, \;\; i = 1, ..., n.$$

and $\hat{y}$ is the posterior mean, i.e., $\hat{y}(\cdot) = \mathrm{E}(y(\cdot) | z(\boldsymbol{s}_1), ..., z(\boldsymbol{s}_n))$.

Connection to Kernel Methods

- You might hear machine learning people talk about "kernel methods" or "kernel machines."
- All this means is that they are modeling $f = \sum c_i K(\cdot, \boldsymbol{x}_i)$.
- They might say, "It allows us to estimate functions over high-dimensional spaces without doing computations in that space." All this means is calculating

$$\hat{y}(\boldsymbol{s}_0) = \Sigma(\boldsymbol{s}_0, S)\Sigma(S, S)^{-1}\boldsymbol{y}$$

  doesn't depend on the dimension of $\boldsymbol{s}_0$.
- They might even say, "This allows us to map our data to an infinite-dimensional space...." You should probably just walk away.

Where are we?

1. Kriging and Isotropy
2. Kriging for Spatio-Temporal Data
3. Other Interpretations of Kriging
4. Wrapping Up

Summary

- The isotropy assumption reduces a high-dimensional problem to a 1-dimensional one.
- Kriging extends in a straightforward way to space-time, but models are a bit disappointing.
- Kriging is essentially an interpolation method.
- Kernel methods (machine learning) are just kriging.

Next Week

- No lectures Monday and Wednesday. (All of us will be away at the Joint Statistical Meetings in Boston.)
- Project presentations during scheduled workshops next Thursday and Friday.
- Good chance to get feedback on your work <u>before</u> you turn it in.
- We'll distribute sign-up sheets. We'll supply refreshments.

Homeworks

- Homework 4a and 4b are now both posted. They are due in two weeks.
- Make sure you download the latest versions of both files. (There were some updates over the weekend.)
- You only need to do one of them. (You are welcome to do both; you must do both if you missed an earlier homework.)
- They are very different but about the same difficulty. Should take less time than Homework 3.
- Homework 4a has an extra credit portion.