

Atividade 1: Análise de dados espaço-temporais para previsão de chuva

Guilherme Ludwig

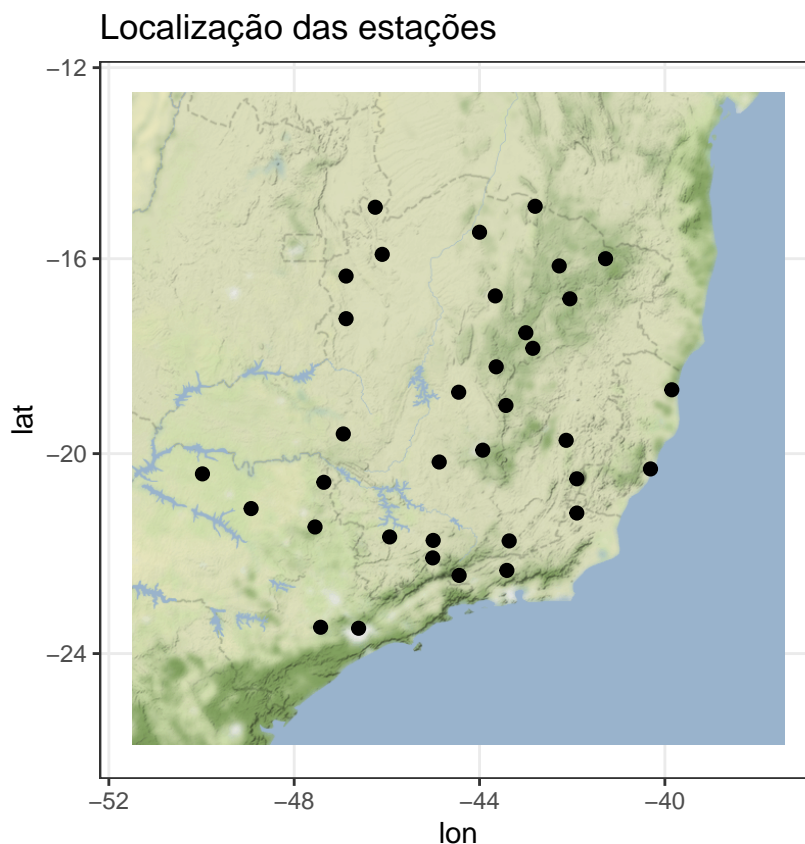
18 de setembro, 2019

Contexto

O objetivo deste projeto é utilizar propriedades da distribuição Normal multivariada para realizar previsões em climatologia. Para tanto, devemos considerar alguma forma paramétrica para Σ , isto é, Σ_{θ} com θ relativamente pequeno

Para mostrar a aplicação imediata desses procedimentos, iremos analisar dados publicados pelo Instituto Nacional de Meteorologia (INMET). O INMET disponibiliza séries temporais obtidas de estações de coleta automática, com medições de temperatura, radiação solar, chuva, etc.; vejam mais a respeito em <http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesAutomaticas> e http://www.inmet.gov.br/portal/css/content/topo_iframe/pdf/Nota_Tecnica-Rede_estacoes_INMET.pdf.

Não há estação climática em Campinas-SP (latitude = -22.9329, longitude = -47.0738).



Objetivo

Nós temos interesse em particular no volume de chuva na região. Os dados de chuva não podem ser utilizados diretamente, pois a sua variabilidade é muito alta. Uma estratégia comum é acumular os dados em períodos

como intervalos de dez dias. Eu extraí e preparei o conjunto de dados para vocês, acumulando a chuva em períodos de 10 dias, e selecionando somente as estações com várias observações em todos os períodos.

```
str(precipitation <- read.csv("precipitacao-verao-2018.csv",
                             stringsAsFactors = FALSE))
```

```
## 'data.frame': 315 obs. of 5 variables:
## $ Cidade      : chr "aracuai-mg" "aracuai-mg" "aracuai-mg" "aracuai-mg" ...
## $ Latitude    : num -16.8 -16.8 -16.8 -16.8 -16.8 ...
## $ Longitude   : num -42 -42 -42 -42 -42 ...
## $ Período     : chr "2018-01-01" "2018-01-11" "2018-01-21" "2018-01-31" ...
## $ Precipitacao: num 25 0 0 555 1.8 ...
```

```
precipitation$Período <- as.Date(precipitation$Período, "%Y-%m-%d")
```

- Organize os dados de chuva em formato vertical. Se t_1, \dots, t_T denotarem os primeiros dias nos intervalos de 10, e s_1, \dots, s_S denotarem as estações, queremos

$$\text{vec}(\mathbf{y}) = \begin{pmatrix} y(t_1, \mathbf{s}_1) \\ y(t_2, \mathbf{s}_1) \\ \vdots \\ y(t_T, \mathbf{s}_1) \\ y(t_1, \mathbf{s}_2) \\ \vdots \\ y(t_T, \mathbf{s}_S) \end{pmatrix}_{ST \times 1}.$$

Observe que $\text{Var}(\text{vec}(\mathbf{y})) = \Sigma_S \otimes \Sigma_T$.

- Nós vamos postular o modelo exponencial para a variância espacial e temporal. Isto é,

$$\Sigma_T = \text{Cov}(y(t_i, \mathbf{s}), y(t_j, \mathbf{s})) = \sigma^2 \exp \left\{ -\frac{|i - j|}{\phi_T} \right\}$$

e

$$\Sigma_S = \text{Cov}(y(t, \mathbf{s}_i), y(t, \mathbf{s}_j)) = \exp \left\{ -\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\phi_S} \right\}.$$

Note que Σ_T é construída com base nos índices (pois os pontos são equidistantes) mas Σ_S é baseada na distância entre as bases.

- A função de log-verossimilhança dos parâmetros $\boldsymbol{\theta} = (\mu, \sigma^2, \phi_T, \phi_S)$ é dada por

$$\log \mathcal{L}(\mu, \sigma^2, \phi_T, \phi_S) = -\frac{ST}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma_S| - \frac{S}{2} \log |\Sigma_T| - \frac{1}{2} (\mathbf{y} - \mu \mathbf{1})^t (\Sigma_S^{-1} \otimes \Sigma_T^{-1}) (\mathbf{y} - \mu \mathbf{1}),$$

você deverá implementá-la como `f(theta)` (onde `theta` corresponde aos quatro parâmetros em $\boldsymbol{\theta}$) e encontrar seu máximo usando `optim` (selecione valores iniciais com cuidado; os parâmetros σ^2 , ϕ_T e ϕ_S devem ser positivos).

- Com $\hat{\boldsymbol{\theta}}$, você poderá construir previsões usando a esperança condicional $\mathbb{E}(y(t_{\text{novo}}, \mathbf{s}_{\text{novo}}) | \mathbf{y}) = \hat{y}(t_{\text{novo}}, \mathbf{s}_{\text{novo}})$, dada por

$$\hat{y}(t_{\text{novo}}, \mathbf{s}_{\text{novo}}) = \hat{\mu} + \text{Cov}_{\hat{\boldsymbol{\theta}}}(y(t_{\text{novo}}, \mathbf{s}_{\text{novo}}), \mathbf{y}) \Sigma_{\hat{\boldsymbol{\theta}}}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}).$$

- Preveja o volume de chuva em Campinas-SP (`latitude` = -22.9329, `longitude` = -47.0738) durante o período dos dados. Isto é, produza um vetor com 9 observações.

Formalidades

- O seu trabalho deverá ser submetido da seguinte maneira: um arquivo nomeado com seu **ra** e **at1**, com extensão **.R** (se você usou R) ou **.py** (se você usou Python). O código deve se comportar como se existisse um arquivo **"precipitacao-verao-2018.csv"** no mesmo diretório. Por exemplo, **123456at1.R** é um nome válido. Eu irei executar esse arquivo com R **CMD BATCH 123456at1.R** ou **python 123456at1.py** (coloque no cabeçalho se precisar de uma versão específica, eu usarei **python3, 3.6.5**).
- O arquivo deve executar sem erro: ler os dados brutos, limpá-los, otimizar a verossimilhança, e produzir a previsão. Não salve nada no disco. Ao final, salve a previsão em formato **.txt**, separado por vírgulas, das 9 previsões para campinas, com o nome igual ao seu **ra** mais **campinas.txt**, por exemplo **123456campinas.txt**.
- A não compreensão das instruções acarretará em penalização da nota, independente do seu código estar correto ou não. Devo reiterar: nomeiem o arquivo corretamente, com a extensão correta. **NÃO** enviem arquivos Rmarkdown com extensão trocada, façam um script simples. **NÃO** salvem as previsões como outro formato senão o especificado, e usem a extensão correta. **NÃO** produzam gráficos que não foram solicitados. **NÃO** usem **setwd("C:\meuUsername\Documents and Settings")** em nenhum lugar do código. Num ambiente profissional é preciso que todos sejam capazes de ler, compreender e seguir instruções simples.

Bom trabalho!