

Laboratorio 3

Óscar Agüero Rodríguez

14/10/2020

Laboratorio 3 Geoestadística

Introducción

La geoestadística es la rama de la estadística que se encarga del análisis de fenómenos espaciales que exhiben un comportamiento estructural. Parte de los mecanismos que están asociados a la geoestadística se encuentra la interpolación, la cual es un proceso donde se utilizan puntos con valores conocidos para estimar valores desconocidos en otros puntos.

A modo de ejemplo, con la interpolación espacial se puede estimar las temperaturas en lugares que no tienen ese dato utilizando lectura de temperatura conocida en estaciones meteorológicas cercanas. A este tipo de superficie interpolada con frecuencia se le llama superficie estadística. Datos de elevación, precipitación, acumulación de nieve, tabla de agua y densidad de población son otros tipos de datos que se puede calcular con la interpolación.

Propiamente, en el documento a tratar se centra en un ejercicio de interpolación para la precipitación en California utilizando puntos con información conocida con los cuales se interpola para obtener la temperatura de los puntos donde no se conoce la cantidad de precipitación.

Respuestas

Pregunta 1: ¿Describa que hace cada chunk de código?

Lo primero que se hace es cargar los datos de precipitación de California, seguido de eso se realiza una proyección de los datos empleado "CRS("+proj=longlat +datum=NAD83"", con lo cual se procede a graficar un mapa con escala de colores según la cantidad de precipitación. En este mapa se observa como en las zonas costeras de california y la zona más al norte es donde se tiene una mayor cantidad de precipitación, no así en la zona más al sur en el mapa.

Después de elaborar el mapa, se realiza una transformación a nivel de latitud y longitud utilizando el sistema de coordenadas conocido como "Teale Albers", esto se realiza para que la interpolación este alineada a los datos.

Una vez que se realiza los ajustes en coordenadas, se procede a calcular el modelo nulo, el cual, es la forma más sencilla para comparar una vez que se tengan otros enfoques. La comparación se realiza utilizando el Error Cuadrático Medio (RMSE).

Posterior a calcular el modelo nulo, se calcula los polígonos de proximidad, el cual es otro método de interpolación espacial, donde se busca el vecino más cercano. Se gráfica un mapa con los resultados de aplicar voronoi. Siguiendo el código, se realiza un mapa aplicando raster considerando que luego de este mapa

rasterizado, se aplica una validación cruzada 5 veces, con la cual se determina la cantidad de precipitación para California, en este caso también se compara por medio del RMSE, el cual para el método de polígonos de proximidad se obtiene un menor RMSE.

Pregunta 2: ¿Cómo se compara el enfoque de polígono de proximidad con el modelo nulo?

La comparación se puede realizar por medio del cálculo del RMSE, el cual para el modelo nulo es de 435.3 y en el caso del modelo utilizando polígonos de proximidad es de 196.7, entendiendo que al utilizar el error cuadrático medio lo que se busca es menor error cuadrático medio, el modelo de polígonos de proximidad es el que da mejores resultados en términos de la interpolación de la variable precipitación.

Pregunta 3: ¿Normalmente no se utilizarían polígonos de proximidad para datos de precipitación para que tipo de datos si los usaría?

La proximidad se utiliza para comparar la cercanía entre dos datos, lo que se puede llamar zona de influencia, al aplicar la interpolación lo que se obtiene es el promedio de esos dos puntos espaciales, este tipo de técnicas son mejor empleadas con datos que permitan delimitar la proximidad entre dos puntos:

- Proximidad a puertos y aeropuertos con respecto a un punto de interés.
- Proximidad de zonas industriales a sectores de residencia o a centro de alta concentración poblacional.
- Proximidad a calles principales desde una ubicación específica, por ejemplo, una escuela.
- Proximidad a fronteras desde los centros industriales.

Pregunta 4: Los raster generados por IDW tienden a tener un artefacto notable, ¿qué es eso?

La distancia Inversa Ponderada (IDW) es el método de interpolación en la que los puntos de muestreo se ponderan durante la interpolación de tal manera que la influencia de un punto en relación con otros disminuye con la distancia desde el punto desconocido que se desea crear. Se puede ver en la aplicación del caso que genera un mapa en el cual se indican menos cantidades de precipitación, en su mayoría alrededor de los 500 ml al observar todo el mapa, con lo que se podría decir que la interpolación anterior aplicada estaba sobre estimando los resultados.

Pregunta 5: Inspeccione los argumentos utilizados para generar un mapa del modelo IDW. ¿Qué nombre le pondría a este método y por qué?

Los argumentos utilizados son similares a los que se utilizan en la interpolación con polígonos de proximidad, con la variante de que en `nmax = 1` y en `set = list(idp = 1)`, en el primer parámetro se está contemplando "nmax" define la proximidad con la observación más cercana que se va a usar para predicción kriging. En el caso de `idp = 1`, se busca el objeto con la distancia más corta.

Este tipo de modelo yo lo llamaría vecino más cercano, similar al otro modelo de polígonos de proximidad, sin embargo, este tiene la particularidad de que solo toma de referencia la unidad más cercana.

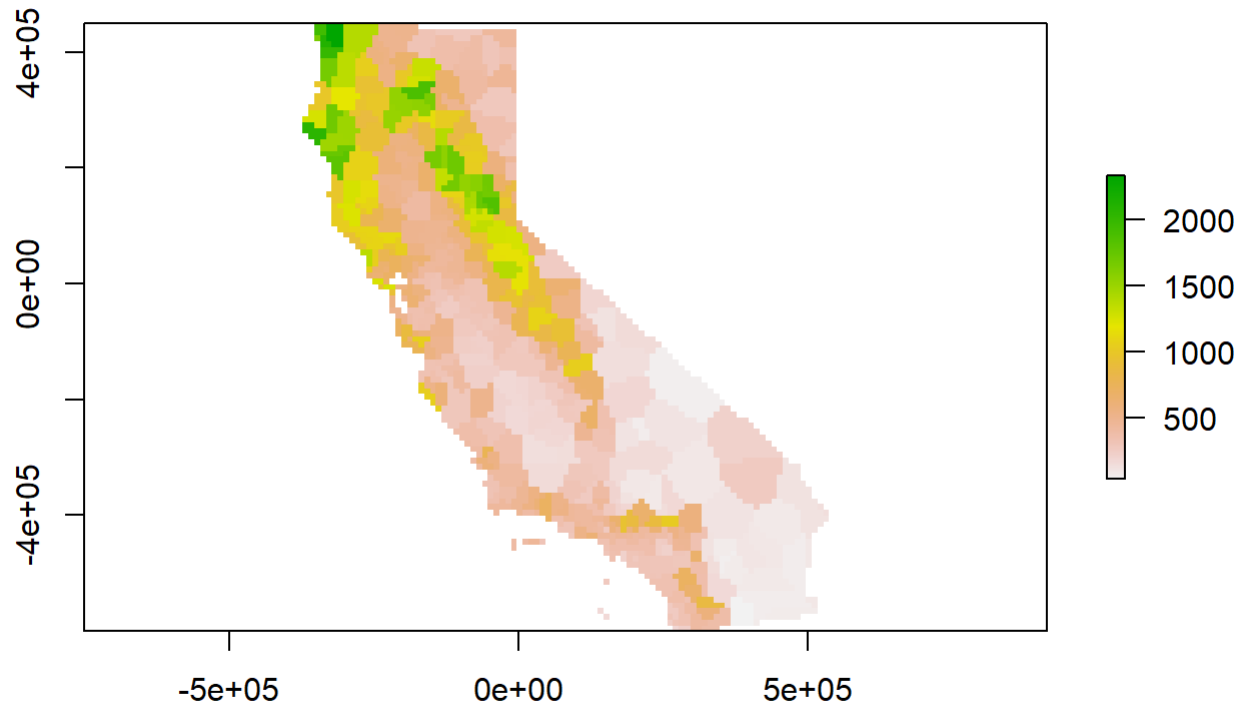
Mapa con el segundo modelo:

```
library(sp)
library(raster)
library(rspatial)
library(gstat)
library(rgdal)
library(deldir)
library(dismo)
library(rgeos)
```

```
d <- sp_data('precipitation')
d$prec <- rowSums(d[, c(6:17)])
dsp <- SpatialPoints(d[,4:3], proj4string=CRS("+proj=longlat +datum=NAD83"))
dsp <- SpatialPointsDataFrame(dsp, d)
CA <- sp_data("counties")
cuts <- c(0,200,300,500,1000,3000)
TA <- CRS("+proj=aea +lat_1=34 +lat_2=40.5 +lat_0=0 +lon_0=-120 +x_0=0 +y_0=-4000000 +datum=NAD83 +units=m +ellps=GRS80 +towgs84=0,0,0")
dta <- spTransform(dsp, TA)
cata <- spTransform(CA, TA)
ca <- aggregate(cata)
v <- voronoi(dta)
vca <- intersect(v, ca)
r <- raster(cata, res=10000)
vr <- rasterize(vca, r, 'prec')
gs2 <- gstat(formula=prec~1, locations=dta, nmax=1, set=list(idp=1))
idw2 <- interpolate(r,gs2)
```

```
## [inverse distance weighted interpolation]
```

```
idwr2 <- mask(idw2,vr)
plot(idwr2)
```



Pregunta 6: ¿Cuál método funciona mejor?

Comparando los métodos por medio de RMSE, se puede ver que entre el IDW óptimo, Kriging ordinario, Spline de placa delgada y el método de conjunto ponderado, el método que funciona mejor es el de conjunto ponderado, ya que este presenta el menor RMSE, el cual es de 7,8580, tan solo un poco menor que el método de Spline el cual tiene un RMSE igual a 7,93007

Pregunta 7: ¿Muestre cual es la mayor diferencia entre IDW y OK?

Se puede ver como los resultados obtenidos por medio de IDW concentran los niveles de contaminación en la zona sur de california, con la particularidad de que abarca una mayor área de todo el mapa en la zona sur, diferente a lo que muestra OK, el cual si bien, también muestra que la concentración de contaminación es mayor al sur de california, muestra que los niveles altos de contaminación se ubican al centro del mapa en el sur del mismo.

Pregunta 8: ¿Muestre dónde está la diferencia entre IDW y OK dentro del límite de confianza del 95% de la predicción OK?

Las diferencias se pueden ver a nivel de los mapas, ya que en la interpolación de los mapas se aprecia que la interpolación con IDW genera una predicción de casos de alta contaminación principalmente en la zona sur de California, distinto a lo que se logra ver utilizando OK. No pude calcular los límites de confianza, pero tendería a pensar que los límites del método IDW son más amplios que los que da el método OK.

Pregunta 9: ¿Puede describir el patrón que se está observando y especular sobre que lo puede estar causando?

Especulo que el patrón observado obedece a una predicción donde los límites de confianza son mayores de ahí que en IDW se tenga una mayor cobertura de contaminación en la zona sur de California, distinto a lo que se observa en el método OK y en TPS.