# Spatial Data and Analysis

# Discussion 5: Review Session

**Felipe González**

UC Berkeley

October $2nd$

# Outline

1. Distances
2. Geographic coordinates
3. Point processes
4. Polylines
5. Networks
6. Fields
7. High dimensional data
8. Matlab

# Distances

- Things you need to know:

  1. How to calculate distances
  2. Use triangulation and trilateration
  3. Time, speed, and velocity

# Calculate distances

- Let $\vec{x_1}$ and $\vec{x_2}$ be two locations in $\mathbb{R}^2$
- Calculate the Euclidean distance between them:

$$D(\vec{x_1}, \vec{x_2}) = \sqrt{(x_1(1) - x_2(1))^2 + (x_1(2) - x_2(2))^2}$$

- You should be able to generalize this concept to $\mathbb{R}^n$

# Triangulation and trilateration

- Triangulation allow us to guess a distance of interest from angles and other distances
- Example on whiteboard

# Triangulation and trilateration

- Triangulation allow us to guess a distance of interest from angles and other distances
- Example on whiteboard
- Trilateration allow us to uniquely guess a location using distances as information
- Example on whiteboard

# Time, speed and velocity

- You need to remember the following relationship:

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

- And be able to perform simple calculations with it

# Midterm question

- Suppose we know the exact location of two landmarks within a city: City Hall and a statue that is 2 km away from City Hall. I tell you to meet me at a location that is 2 km from City Hall and 1.5 km from the statue. Do you have enough information to know where to meet me?

  1. Yes
  2. No

# Geographic coordinates

- Things you need to know:

  1. Latitude:
     - specifies north-south position
     - ranges from 0° at the equator to 90° (north or south)
     - 1 degree = 111 km
  2. Longitude
     - specifies west-east position
     - ranges from 0° at the Greenwich Meridian to 180° (west or east)
     - 1 degree = 111 km $\times \cos(\theta)$

# Midterm question

- A plane travels from 0°N Latitude and 0°E Longitude to 0°N Latitude and 20°E Longitude in 6 hours. What is the average speed of the plane?

    1. 270 km/hr
    2. 330 km/hr
    3. 185 km/hr
    4. 33 km/hr
    5. none of the above

# Midterm question

- If we approximate a field covering the surface of the Earth using a 1° latitude by 1° longitude raster, how many grid cells will our field have?

  1. 8,100
  2. 16,200
  3. 32,400
  4. 64,800
  5. 129,600
  6. none of the above

# Point processes

- Things you need to know:

  1. Intensity of point processes
  2. $k$-function
  3. Calculate centroids
  4. Interpolate using nearest neighbor and other weights

# Point processes

- Intensity is simply the number of events over area:

$$\hat{\lambda} \;=\; \frac{Z}{A}$$

- $k$-function measures the number of events in a certain area from a certain point:

$$\hat{k}(h) \;=\; \frac{1}{\hat{\lambda}} \sum_i \sum_j \frac{1[d(\vec{s}_i, \vec{s}_j) < h]}{Z}$$

# Inverse distance weighting

- Estimates using inverse distance weighting:

$$\hat{Z}(x) \;=\; \frac{1}{\sum_i d(x, x_i)^{-k}} \left( Z_1 d(x, x_i)^{-k} + \cdots + Z_N d(x, x_N)^{-k} \right)$$

- Think about it as an average of values close by
- Note that the right hand side is simply a weighted average, a generalization of an average

# Midterm question

- Last year there were 6,743 crimes in Berkeley (area 46 sq. km), 28,085 crimes in Oakland (202 sq. km) and 39,244 crimes in San Francisco (600 sq. km). Which city has the highest average spatial density of crime events?

    1. Berkeley
    2. Oakland
    3. San Francisco
    4. not enough information

# Midterm question

- What can we say about two point processes $A$ and $B$ that have the following $k$-functions:

$$k_B(h) \ = \ k_A(h) \ = \ \pi \times h^2$$

1. $A$ is homogenous
2. $B$ is homogenous
3. $A$ and $B$ are probably Poisson Point Processes
4. $A$ and $B$ are probably not Poisson Point Processes
5. $\lambda_A(\vec{\c{c}}) = \lambda_B(\vec{\c{c}})$
6. none of the above

# Midterm question

- Given the set of points $\mathcal{S} = \{[2,3];[4,9];[-3,0]\}$, the centroid of $\mathcal{S}$ is:

    1. $[1,4]$
    2. $[1.5,6]$
    3. $[3,6]$
    4. $[2,3]$
    5. none of the above

# Midterm question

- You're given the following data with locations and income:

$$\tilde{H} = \begin{Bmatrix} \vec{\underset{\cdot}{c}}_1 = [6,4], & Y_1 = \$24,000 \\ \vec{\underset{\cdot}{c}}_2 = [2,7], & Y_2 = \$32,000 \\ \vec{\underset{\cdot}{c}}_3 = [1,3], & Y_3 = \$30,000 \end{Bmatrix}$$

Use (1) nearest-neighbor, and (2) inverse distance weighting interpolation, to estimate the income of a household located in $\vec{\underset{\cdot}{c}} = [3,4]$. Which one do you prefer?

# Polylines

- Things you need to know:

  1. Definitions:
     - lines
     - polygon
     - polylines
     - multiple polygons
  2. Concepts:
     - concavity
     - convexity

# Definitions

- Line: $\ell = \{\vec{x}, \vec{y}\}$, where $\vec{x} = [x_1, x_2]'$ and $\vec{y} = [y_1, y_2]'$
- Polyline: $L = \{\vec{x}, \vec{y}\}$, where $\vec{x} = [x_1, \ldots, x_N]'$ and $\vec{y} = [y_1, \ldots, y_N]'$
- A polygon $P$ is a polyline where $x_1 = x_N$ and $y_1 = y_N$
- A multiple polygon is a set of polygons $H = \{P_1, P_2, P_3\}$
- A polygon is convex if a linear combination of any two points remains inside the polygon. Concave $= \neg$ (Convex)

# Midterm question

- A 5-pointed star is a convex shape

  1. true
  2. false

# Midterm question

- Suppose you are given this structure in Matlab:

```
A.x = [0 2 1 0 NaN 0 -1 0 1 0 NaN 3 4 4 3 3]
A.y = [0 -1 1 0 NaN 4 3 2 3 4 NaN 2 2 3 3 2]
```

Are each of the polygons individually convex?

1. Yes
2. No
3. not enough information

# Networks

- Things you need to know:
  1. Definitions
     - network
     - connectivity matrix
  2. How to interpret a connectivity matrix
  3. What is a stimulus

# Definitions

- A network is a set of vertices and edges with a structure composed by attributes and a connectivity matrix:

$$\Gamma \;=\; \left\{\vec{x}, \vec{y}, \vec{A}, L\right\}$$

- Let $N$ be the number of vertices. Then, the connectivity matrix $L$ has dimension $N \times N$, with $L(i,j) = 1$ if $i$ and $j$ connected and $L(i,j) = 0$ otherwise

- A stimulus is a signal transmitted through the network. $L \times \vec{S}$ receives the signal

# Midterm question

- Individuals with connectivity matrix: $L = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$

  Which statement are true?

  1. there are 10 individuals in the network
  2. individual 3 and 4 are connected
  3. individual 1 and 4 are not connected
  4. individual 5 is considered ‘connected’ to herself
  5. none of the above

# Midterm question

- Connectivity matrix for ports: $L = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$
- If a crate of coffee begins at location 1, is it possible to go to 5 using existing routes?
  1. Yes
  2. No

# Fields

- Things you need to know:
  1. Definitions:
     - field
     - raster
     - contour
- What is map algebra

# Fields

- A field is a function over space, i.e., an attribute at every location
- A raster is the approximation of a field using a grid where points have a uniform value in each grid of area $\delta_x \cdot \delta_y$
- A contour is a polyline where the field value at points on the line is constant

# Map algebra

- Map algebra is simply the computation of functions of fields on a point-wise basis, i.e., computed at each point separately
- Masking is one type of map algebra in which one of the fields, the mask, is binary

# Midterm question

- What are examples of a field?

  1. temperature
  2. the risk of a car accident occurring
  3. all the rivers in the world
  4. your address
  5. travel time to Times Square, NYC
  6. the highest elevation in the country
  7. none of the above

# Midterm question

- Which are examples of a static field?

1. the quality of soil across Wyoming
2. the length of each river segment in Brazil
3. the 3-dimensional centroid of each room in GSPP
4. travel time to Ghirardelli Square, SF
5. the population of all counties in China
6. the average crime rate
7. none of the above

# High dimensional data

- A vector field is a vector of data $\vec{f}$ at each location $\vec{c}$
- A dynamic field is a vector of fields in which elements in the vector represent sequential observations
- A dynamic vector field as multiple variables that vary over time at every point

# Midterm question

- We collect temperature and rainfall data for all of Africa every day for a year. This data set represents a:

  1. dynamic vector field
  2. static field
  3. gradient
  4. point process
  5. delimited shapefile
  6. kernel estimate
  7. none of the above

# Matlab

- Things you need to know:
    1. Concepts we used in Labs
    2. Be ready to think in code terms
    3. Be ready to explain rough steps on how to code something

# Midterm question

- A mask is useful for:
  1. measuring the bandwidth of a kernel
  2. computing the $k$-function for a Poison point process
  3. isolating raster data in a region of interest
  4. none of the above

# Midterm question

- In Matlab, a buffer is:
  1. field
  2. raster
  3. polygon
  4. topology
  5. point process
  6. moving window

The End