CSci 5715, Fall 19: Homework 4

Due on 11/26 Before Class

Table of Participation

Question ID	Answer drafted by	Answer reviewed by
1	Alex	Majid
2	Alex	Majid
3	Alex	Majid
4	Majid	Alex
5	Majid	Alex

Question 1. Figure 1 shows the occurrences of three types of events represented by red circle, blue triangle, and yellow square. It further shows, two different types of spatial partitioning (A, B). Assume that events co-occur if the distance between their closest point is at most 1.5 units.

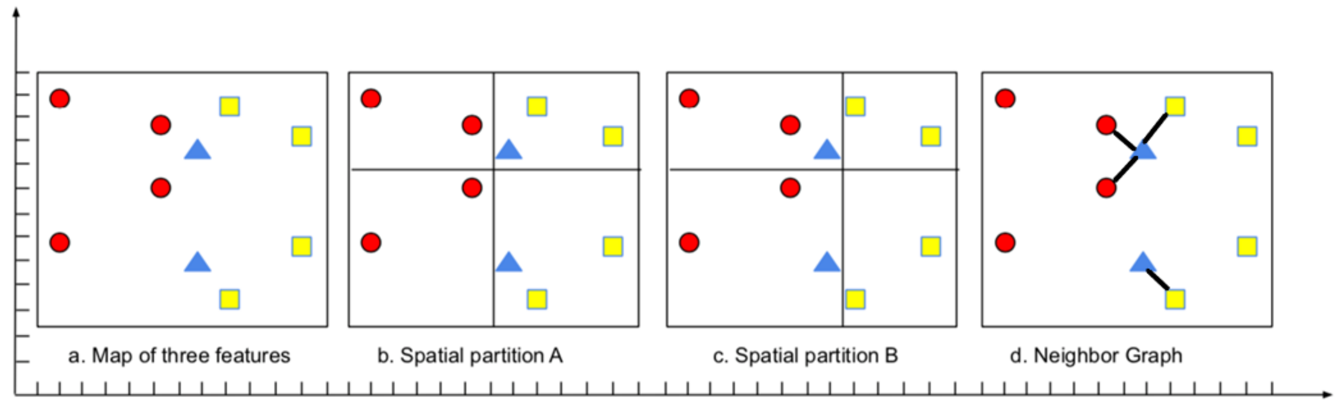


Figure 1. Neighbor relationship vs. Space Partitioning.

Q1a). Draw the neighborhood graph for the events shown in Figure 1. You may use the space provided in **1d.**

Q1b).. Fill out Table 1 to show the values of Pearson's Correlation (for partition A, and partition B), Support (partition A, and partition B), Ripley's Cross-K and Participation Index. For calculating Pearson's Correlation refer to **Appendix B**.

Table 1. Results of co-occurrence measures

$for\ x = red, y = blue, z = yellow;\ \ n = 4\ (partitions);\ \ \sum_{Part.A} x = 4, \sum_{Part.A} y = 2, \sum_{Part.A} z = 4;\ \ \sum_{Part.A} x * y = 0, \sum_{Part.A} y * z = 4, \sum_{Part.A} x * z = 0;\ \ \sum_{Part.A} x^2 = 8, \sum_{Part.A} y^2 = 2, \sum_{Part.A} z^2 = 8$

$\sum_{Part.B} x = 4, \sum_{Part.B} y = 2, \sum_{Part.B} z = 4;\ \ \sum_{Part.B} x * y = 4, \sum_{Part.B} y * z = 0, \sum_{Part.B} x * z = 0;\ \ \sum_{Part.B} x^2 = 8, \sum_{Part.A} y^2 = 2, \sum_{Part.A} z^2 = 8$

$$PC_A(x,y) = \frac{(4*0)-4*2}{\sqrt{4*8-(4)^2}*\sqrt{4*2-(2)^2}} = \frac{-8}{\sqrt{16}*\sqrt{4}} = \frac{-8}{4*2} = -1$$

$$PC_A(y,z) = \frac{(4*4)-2*4}{\sqrt{4*2-(2)^2}*\sqrt{4*8-(4)^2}} = \frac{8}{\sqrt{4}*\sqrt{16}} = \frac{8}{2*4} = 1$$

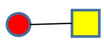
$$PC_A(x,z) = \frac{(4*0)-4*4}{\sqrt{4*8-(4)^2}*\sqrt{4*8-(4)^2}} = \frac{-16}{\sqrt{16}*\sqrt{16}} = \frac{-16}{4*4} = -1$$

$$PC_B(x,y) = \frac{(4*4)-4*2}{\sqrt{4*8-(4)^2}*\sqrt{4*2-(2)^2}} = \frac{8}{\sqrt{16}*\sqrt{4}} = \frac{8}{4*2} = 1$$

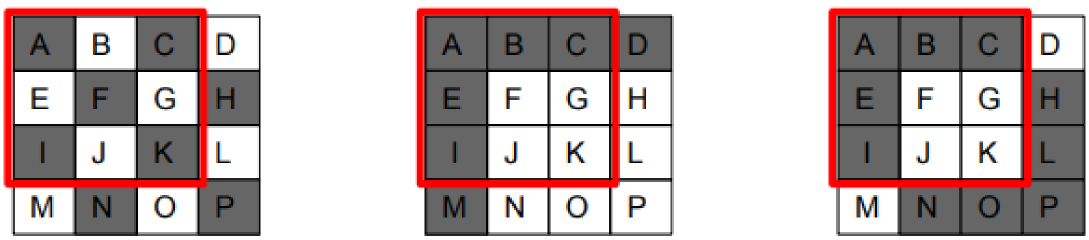
$$PC_B(y,z) = \frac{(4*0)-2*4}{\sqrt{4*2-(2)^2}*\sqrt{4*8-(4)^2}} = \frac{-8}{\sqrt{4}*\sqrt{16}} = \frac{-8}{2*4} = -1$$

$$PC_B(x,z) = \frac{(4*0)-4*4}{\sqrt{4*8-(4)^2}*\sqrt{4*8-(4)^2}} = \frac{-16}{\sqrt{16}*\sqrt{16}} = \frac{-16}{4*4} = -1$$

$Participation\ Ratio(x,y) = 1/2$
$Participation\ Ratio(y,x) =$ ½
$Participation\ Ratio(y,z) = 1$
$Participation\ Ratio(z,y) =$ ½
$Participation\ Ratio(x,z) = 0$
$Participation\ Ratio(z,x) = 0$

	Pearson's Correlation (Partition A)	Pearson's Correlation (Partition B)	Support (Partition A)	Support (Partition B)	Ripley's Cross-K	Participation Index
	**-1**	**1**	**0**	**0.5**	**2/8 = 0.25**	**½**
	**1**	**-1**	**0.5**	**0**	**2/8 = 0.25**	**½**
	**-1**	**-1**	**0**	**0**	**0/16 = 0**	**0**

Q2) Consider a 4 X 4 square with 16 cells, in a raster dataset as shown below in **Figure 2.** Considering the three different cases/patterns below to calculate the spatial auto-correlation for each of the patterns. For the computation purposes assume that dark boxes have value 1, and white boxes have value 0. Further, assume that the neighborhood of a cell consists of the cells **sharing an edge** with the cell. For example, B and E are neighbors of A.



(a) **(b)** **(c)**

Figure 2. Raster dataset.

Q2 a) Moran's I can be computed as follows I = $\frac{zWz^T}{zz^T}$. W_{ij} is defined as 1 if cells i, and j are neighbors, and 0 otherwise. Show the neighborhood matrix W for the bounded region. (**Note.** The W matrix would be same for all the datasets.)

	A	B	C	E	F	G	I	J	K
A	0	1	0	1	0	0	0	0	0
B	1	0	1	0	1	0	0	0	0
C	0	1	0	0	0	1	0	0	0
E	1	0	0	0	1	0	1	0	0
F	0	1	0	1	0	1	0	1	0
G	0	0	1	0	1	0	0	0	1
I	0	0	0	1	0	0	0	1	0
J	0	0	0	0	1	0	1	0	1
K	0	0	0	0	0	1	0	1	0

Row-normalized:

	A	B	C	E	F	G	I	J	K
A	0	.5	0	.5	0	0	0	0	0
B	.33	0	.33	0	.33	0	0	0	0
C	0	.5	0	0	0	.5	0	0	0
E	.33	0	0	0	.33	0	.33	0	0
F	0	.25	0	.25	0	.25	0	.25	0
G	0	0	.33	0	.33	0	0	0	.33
I	0	0	0	.5	0	0	0	.5	0
J	0	0	0	0	.33	0	.33	0	.33
K	0	0	0	0	0	.5	0	.5	0

Q2 b) Each element in z matrix is defined as $z_{ij} = x_{ij} - \text{mean}(X)$, where mean(X) is the mean of all values. Compute the z matrix for each dataset.

Question: should we consider dataset "X" to be the bounded region, or the entire matrix? If it is the bounded region, then b and c are the same… I will assume entire matrix (this also gives us a Moran's I between [-1,1]

$$mean(X_a) = \frac{8}{16} = 0.5, \quad mean(X_b) = \frac{7}{16} = 0.4375, \quad mean(X_c) = \frac{10}{16} = 0.625$$

	A	B	C	E	F	G	I	J	K
z(a)	0.5,	-0.5,	0.5,	-0.5,	0.5,	-0.5,	0.5,	-0.5,	0.5
z(b)	0.5625,	0.5625,	0.5625,	0.5625,	-0.4375	-0.4375	0.5625,	-0.4375	-0.4375
z(c)	0.375,	0.375,	0.375,	0.375,	-0.625,	-0.625,	0.375,	-0.625,	-0.625

Q2 c) Appendix A shows a small example to compute Moran's I value for the 2X2 square with 4 cells. Use the example, to derive the Moran's I values for the 3X3 square with 9 cells (marked in red) and fill the table below. For the computation you can use MATLAB, or Octave. **Note.** Octave is pre-installed in your lab accounts.

Note, I am doing this on the row-normalized matrix.

a.	b.	c.
-1	0.382	0.4

Q3) Suppose a satellite image classification task at hand is performed with a combination of feature engineering using simplified Normalized Difference Vegetation Index (NDVI) and a decision tree classifier. The Red band (R) and Near-Infrared band (NIR) values for different pixels are shown in Figure 3 (a) and (b).

300	280	270
310	250	260
325	230	285

(a) Near-Infrared band

200	200	190
190	210	200
235	140	215

(b) Red band

Figure 3. Input image bands.

Question (a): The simplified NDVI is computed as: $\text{NDVI}_{sim} = NIR - R$. Consider an Indicator function I(NDVI_{sim}) defined for a pixel p:

$$I(\text{NDVI}_{sim}(p)) = \begin{cases} 1; \ if \ \text{NDVI}_{sim}(p) > 50 \\ -1; \ if \ \text{NDVI}_{sim}(p) \leq 50 \end{cases}$$

Fill out the following matrices to show the NDVI_{sim} of the image and the results after applying the Indicator function.

100	80	80
120	40	60
90	90	70

(a) NDVI_{sim}

1	1	1
1	-1	1
1	1	1

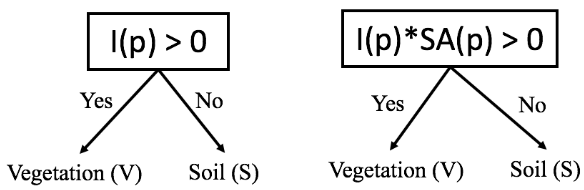
(b) I(NDVI_{sim})

Question (b): Two pixels are neighbors if they are not disjoint, i.e., their boundaries **touch** or **share** at a point. For example, the pixel at the center of the given image has 8 neighbors. Given a pixel **p** in the $NDVI_{sim}$ result image, denote the collection of its neighbors (excluding itself) as **N(p)**. Consider a spatial auto-correlation function of a pixel **p**: $SA(p) = \sum_{q \in N(p)} I(p) * I(q)$, where * denotes multiplication of numbers. For example, the top-left pixel has a SA value is 1 as given below. Show the results of SA(p) for other pixels in the image:

1	3	1
3	-8	3
1	3	1

(a) SA(p)

Question (c): Figure 4(a) shows a traditional decision tree (DT) created for this classification task, and Figure 4(b) shows a spatial decision tree. For example, the top-left pixel (TLP) is classified as vegetation by both decision trees since I(TLP) = 1 and I(TLP)*SA(TLP) = 1. Show the classification results of applying the two decision trees.



(a) Traditional decision tree. (b) Spatial decision tree.

Figure 4. Decision tree(s)

Results:

V	V	V
V	S	V
V	V	V

(a) Traditional Decision Tree

V	V	V
V	V	V
V	V	V

(b) Spatial Decision Tree

Question 4. A ski athlete wants to estimate his position P through 3D triangulation. The athlete uses the following knowledge to estimate the position. **1.** Positional knowledge of two consecutive chairlift stations (L_1, L_2), and **2.** A theodolite [1] to estimate the angle (α, β) between each of the stations. 3D triangulation uses two 2D triangulation to compute the position. The following figures are from slide #27 on positioning, that shows the formulation of two 2D triangulations. (**Note:** $\sin^{-1}(0.78)$ = 0.89 radians, or 51.26 degrees. cos(0.89) = 0.63).

There are two construction(s).

1. As shown in Figure 4 ~~b~~ c. **PC** is the line perpendicular to line L_1L_2.

2. As shown in Figure 4 ~~c~~ d. (L1', L2', C') are the xy plane projections of (L1, L2, C).

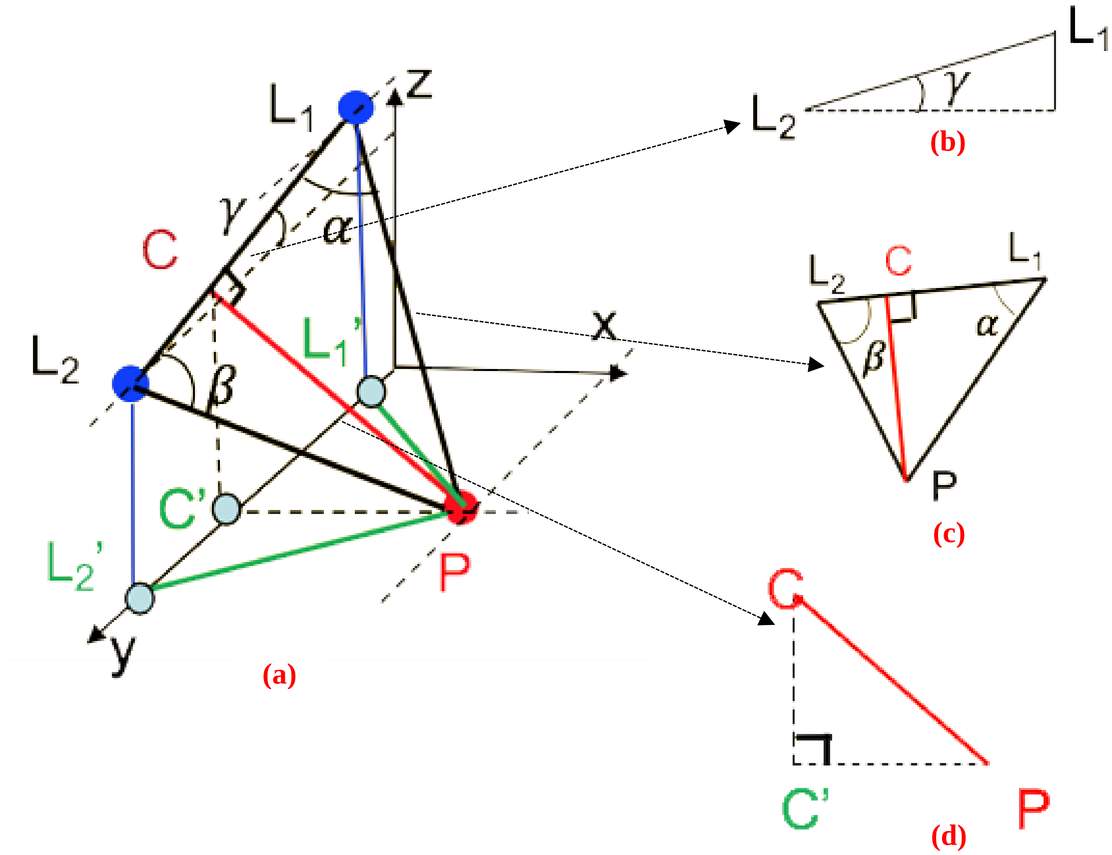


Figure 4. 3D Triangulation using two 2D triangulation.

Using the following numerical values estimate the position of the athlete.

Coordinates of L1 (x, y, z)	Coordinates of L2 (x, y, z)	Angle (L_1P, L_1L_2)	Angle (L_2P, L_1L_2)
(0, 12, 100)	(0, 72, 25)	60°	60°

First calculating the distance between P, and C as following:

$$\boldsymbol{d(P,C)=\frac{\sqrt{(72-12)^2+\ (100-25)^2}}{cot\ a+cot\ b}\ =\frac{96.07}{1.15}=\ 83.51}.$$

Then in the following we calculate the distance between d (L1, L2) as following:

$$\boldsymbol{d(L1,L2)=d(P,C)*(cota+cot\ b)=96.05}$$

Then we calculate the distance between d (C, C') as the following:

$$\boldsymbol{25+(100-25)*\frac{cot\ 60}{cot\ 60+cot\ 60}=62.5}$$

Calculating the Y degree as the following:

$$\boldsymbol{sin^{-1}=\frac{100-25}{d(l1,l2)}=51.34}$$

Calculating the x and y coordinate for point P as the following:

P_x = $\boldsymbol{d\ (P,C')=\ \sqrt{d(p,c)^2-d(c,c')^2}=54.88}$

Also, we will have the following:

d (L1, C) = d(P,C)cot a = 48.02

Finally:

P_y = 12 + d (L1, C) cos y = 12 * 48.02 * cos 51.34 = 42.

Question 5.
Three sound recorders in city streets recorded sound of a gunshot at 4:05:29 PM, 4:05:30 PM, and 4:05:31 PM respectively. According to the flash light of the gunshot detected by recorders, this gunshot occurred at 4:05:28.500 PM. The locations of these three recorders are (0, 0), (0, 400 meters), and (990.1690 meters, 0), respectively. The speed of sound is 340 meters / second. (Assume that all the recorders and the gunshot are on a two-dimensional plane).

Q5a) Determine the exact position of the gunshot. Justify the answer by showing the intermediate calculations.

Based on GPS, we will have the following equation:

$$x^2 + y^2 = 170^2$$
$$x^2 + (y - 400)^2 = 510^2$$
$$x - 990.1690)^2 + y^2 = 850^2$$

Then solving equation for x, and y we will have the following:

(164.5,44.5)

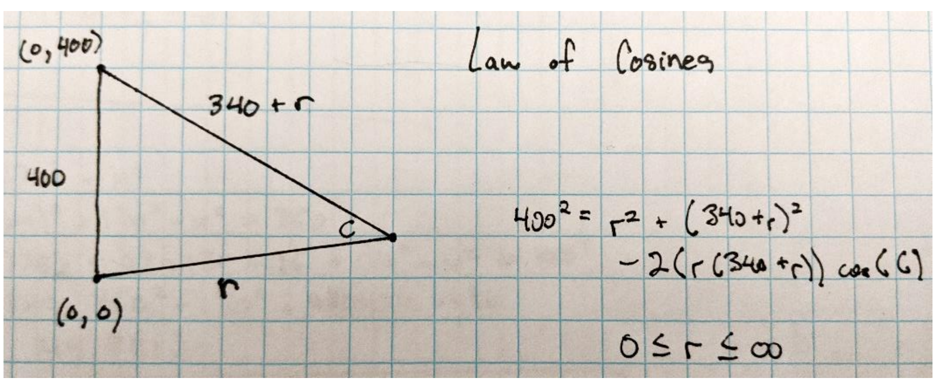
Q5b) Now suppose we do NOT have the third sound recorder at (990.1690 meters, 0). Can we still determine the exact position of the gun shot? Justify your answer.

No, since we need to deal with two different possibilities.

Q5c) Now suppose we do NOT have the third sound recorder at (990.1690 meters, 0). And, we do NOT know the time when the gunshot happened. Give all the possible locations where this gunshot happened.

Given the location of the first two sound recorders, the times of the gunshot, and the time elapsed between when each recorder picked up the shot, we can write a formula for the distance between the recorders.

$x^2 + (y - 400)^2 - x^2 - y^2 = 340 * 1$



If we knew the exact time of the gunshot, we could solve for r and thus for (x,y)

Appendix A.

A	B
C	D

W

	A	B	C	D
A	0	1	1	0
B	1	0	0	1
C	1	0	0	1
D	0	1	1	0

Row Normalized W

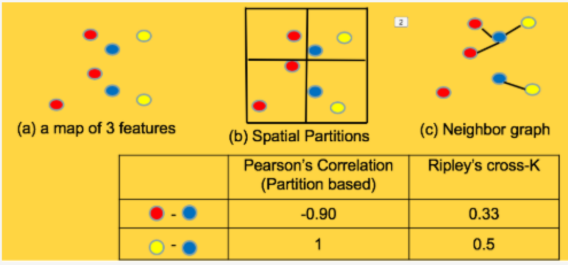
	A	B	C	D
A	0	0.5	0.5	0
B	0.5	0	0	0.5
C	0.5	0	0	0.5
D	0	0.5	0.5	0

z^T

0.5
-0.5
-0.5
0.5

```
octave:1> M = [0 0.5 0.5 0; 0.5 0 0 0.5; 0.5 0 0 0.5; 0 0.5 0.5 0]
M =

   0.00000   0.50000   0.50000   0.00000
   0.50000   0.00000   0.00000   0.50000
   0.50000   0.00000   0.00000   0.50000
   0.00000   0.50000   0.50000   0.00000

octave:2> Z = [0.5 -0.5 -0.5 0.5]
Z =

   0.50000  -0.50000  -0.50000   0.50000

octave:3> (Z*M*Z') / (Z*Z')
ans = -1
octave:4>
```

Appendix B.



Let's take the sample dataset in the first figure part (b) as an example. The data we have is

	red	blue	yellow
1.	1	0	0
2.	0	1	1
3.	2	0	0
4.	0	1	1

According to the Pearson correlation coefficient formula (https://en.wikipedia.org/wiki/Pearson_correlation_coefficient), let's say x is red and y is blue, then the x-y correlation can be calculated.

Calculation for Pearson correlation between red (x), and blue (y),

	x.	y.	x*y	x^2	y^2
1.	1	0	0	1	0
2.	0	1	0	0	1
3.	2	0	0	4	0
4.	0	1	0	0	1
∑	3	2	0	5	2

$$PC(x,y) = \frac{n*\Sigma x*y - \Sigma x*\Sigma y}{\sqrt{n*\Sigma x^2-(\Sigma x)^2}*\sqrt{n*\Sigma y^2-(\Sigma y)^2}}$$

$$PC(x,y) = \frac{0*4-6}{\sqrt{4*5-9}\sqrt{4*2-4}} \approx -0.9$$

Support for red-blue = #partitions where red-blue together / # total partitions = 0 / 4 = 0

Support for blue-yellow = #partitions where yellow-blue together / # total partitions = 2 / 4 = 0.5