Means & Medians to Machine Learning: Spatial Statistics Basics and Innovations

Lauren Bennett
Jenora D’Acosta
Flora Vale

esriurl.com/spatialstats
What are Spatial Statistics?
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.
coincidence
closest point
connectivity
area
proximity
orientation
length
direction
Spreadsheets

Data or Information?
Maps

Data or Information?
When you look at a spreadsheet...
You ask for more

- Mean
- Standard Deviations
- Min and Max
- ...

![Table and Graphs]
Same goes for maps!
We can do more
Means and Medians
summarizing spatial distributions

Machine Learning
clustering methods
Means and Medians
summarizing spatial distributions
Central Feature

identifies the most centrally located feature in a point, line, or polygon feature class
Mean Center

identifies the geographic center (or the center of concentration) for a set of features
(14,14)
(13,12)
(25,24)
(24,16)
(22,23)
(18,12)
(12,12)
(14,8)
(9,18)

mean = (17,15)
Median Center identifies the location that minimizes overall Euclidean distance to the features in a dataset.
median = (14, 14)
Mean vs Median?
Linear Directional Mean identifies the mean direction, length, and geographic center for a set of lines.
Directional Distribution
(Standard Deviational Ellipse)

creates standard deviational ellipses to summarize the spatial characteristics of geographic features: central tendency, dispersion, and directional trends
Demo
Machine Learning

clustering methods
Density-based Clustering
finds clusters based on feature locations
DBSCAN – defined distance
HDBSCAN – self adjusting
OPTICS – multi-scale
DBSCAN – defined distance
DBSCAN – defined distance
DBSCAN – defined distance
DBSCAN – defined distance

- **core distance**
- **search distance**
DBSCAN – defined distance

search distance

core distance

core distance
DBSCAN – defined distance

search distance

core distance

search distance

core distance
OPTICS – multi-scale
OPTICS – multi-scale
OPTICS – multi-scale
OPTICS – multi-scale
OPTICS – multi-scale
<table>
<thead>
<tr>
<th><strong>DBSCAN</strong></th>
<th><strong>HDBSCAN</strong></th>
<th><strong>OPTICS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses fixed search distance</td>
<td>Uses range of search distances to find clusters of varying densities</td>
<td>Uses neighbor distances to create reachability plot</td>
</tr>
<tr>
<td>Clusters of similar densities</td>
<td>Data driven, requires least user input</td>
<td>Most flexibility for fine tuning</td>
</tr>
<tr>
<td>Fast</td>
<td></td>
<td>Can be computationally intensive</td>
</tr>
</tbody>
</table>
Demo
Multivariate Clustering

finds clusters based on feature attributes
K Means
K Means

2 groups
K Means

2 groups

3 groups
K Means

2 groups  3 groups  4 groups
Eligible Uninsured Americans

- % Below 138 FPL
- % Uninsured
- % No High School
- % Latino
Spatially Constrained Multivariate Clustering

finds clusters based on feature attributes and proximity
WHAT'VE YOU BEEN UP TO?
DOING TONS OF
MATH FOR MY THESIS.
CAN YOU EXPLAIN IT LIKE I'M FIVE?
"OH MY GOD, WHERE ARE YOUR PARENTS?"
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Minimum Spanning Tree
Crime in Chicago

- Median Income
- HS Dropout Rate
- Unemployment
- Crime Count
Demo
Build Balanced Zones

creates spatially contiguous zones in your study area using a genetic growth algorithm based on criteria that you specify
Criteria

Zone building

• Attribute target
• Number of zones and attribute target
• Number of zones
Criteria

Zone selection

• Equal area
• Compactness
• Equal number of features
• Attribute to consider
fitness score
9.14
top 50% move on to next generation
top 50% move on to next generation
top 50% move on to next generation
top 50% move on to next generation
and crossover to create new offspring
top 50% move on to next generation
and crossover to create new offspring
top 50% move on to next generation and crossover to create new offspring
then the next fittest 50% moves on and crosses over to create the next generation
Convergence

Fitness Score

Generation
Demo
esriurl.com/spatialstats