GeoMeanShift

Title GeoMeanShift

Description

This scripting tool will take in an input feature class and feed each features centroid coordinates into Scikit-Learns Mean Shift Implemenation. This scripting tool will add the resulting Mean Shift Labels directly to the input feature class. Mean Shift clustering only requires a bandwidth (search radius) to determine clusters of features. Mean shift is a hill climbing algorithm which involves shifting features iteratively to a higher density region (based on a kernel density estimation of the appropriate bandwidth) until convergence.

Source: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html#sklearn.cluster.MeanShift

Wiki: https://en.wikipedia.org/wiki/Mean_shift

Usage

This tool will take points and feed the X-Y coordinates of the current projection as a two dimensional array to be clusted by scikit-learns implementation of Mean Shift. Mean shift is a hill climbing algorithm which involves shifting features iteratively to a higher density region (based on a kernel density estimation of the appropriate bandwidth) until convergence.

Source: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html#sklearn.cluster.MeanShift

Wiki: https://en.wikipedia.org/wiki/Mean_shift

Syntax

GeoMeanShift (Input_Feature_Class, {Search_Radius}, Use_Estimated_Search_Radius, {Cluster_All_Points}, {Ouput_Centroid_Feature_Class}, {Weight_Field})

Parameter	Explanation	Data Type
Input_Feature_Class	Dialog Reference This is the input feature class that will be clustered using Mean Shift and its labels will be added to the feature class fields. Noise is a -1 and each sequential number after is a cluster. If Cluster All Point is is true, there will be no -1 Noise labels. There is no python reference for this parameter.	Feature Layer
Search_Radius (Optional)	Dialog Reference This is the bandwidth that will be used by the kernel density estimation function that will be used to generate the "hills" the points will climb to reach a new cluster. The units of this parameter are in the units of the current projection. There is no python reference for this parameter.	Double
Use_Estimated_Search_Radius	Dialog Reference If True, this tool will used a bandwidth determiend by Scikit Learns estimate band width function. Link: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.estimate_bandwidth.html There is no python reference for this parameter.	Boolean
Cluster_All_Points (Optional)	Dialog Reference If True, true, then all points are clustered, even those "orphans" or noise that are not within any kernel. Orphans are assigned to the nearest kernel. If false,	Boolean

	then orphans are given cluster label -1. There is no python reference for this parameter.	
Ouput_Centroid_Feature_Class (Optional)	Dialog Reference This optional output feature class consists of the centroids of the clusters computed by scikit-learn. The output feature class with be in the same projection as the input and will have fields that denote the cluster label and number of points associated with that centroid. There is no python reference for this parameter.	Feature Class
Weight_Field (Optional)	Dialog Reference This is the number of samples each record represents. For example, if there are 5 entities represented by one point (collision database records for example), then this will weight the flat kernel used to cluster the data based on that field. Null values will be treated a a value of weight 1. Python Reference This function makes use of the np.repeat() function to weight the samples. Might increase run time slightly, but only for cases that make use of a weight field.	Field

Code Samples

There are no code samples for this tool.

Tags

Mean Shift, KDE, Scikit-Learn, Clustering

Credits

David Wasserman

Use limitations

There are no access and use limitations for this item.

You are currently using the Item Description metadata style. Change your metadata style in the Options dialog box to see additional metadata content.