Machine Learning in ArcGIS

By Lauren Bennett, Esri Spatial Analyst

Esri's continued advancements in data storage and parallel and distributed computing make solving problems at the intersection of machine learning (ML) and GIS increasingly possible.

ML refers to a set of data-driven algorithms and techniques that automate the prediction, classification, and clustering of data. ML can be computationally intensive and often involves large and complex data. It can play a critical role in spatial problem-solving in a wide range of application areas from multivariate prediction to image classification to spatial pattern detection.

In addition to traditional ML techniques, ArcGIS also has a subset of ML techniques that are inherently spatial. Spatial methods that incorporate some notion of geography directly into computation can lead to deeper understanding. The spatial component often takes the form of some measure of shape, density, contiguity, spatial distribution, or proximity. Both traditional and inherently spatial ML can play an important role in solving spatial problems. ArcGIS supports the use of ML in prediction, classification, and clustering.

Prediction uses the known to estimate the unknown. ArcGIS includes regression and interpolation techniques that can be used for performing prediction analysis. ArcGIS has tools for empirical Bayesian kriging (EBK), areal interpolation, EBK regression prediction, ordinary least squares (OLS) regression, OLS exploratory regression, and geographically weighted regression (GWR). These tools can be used for tasks like estimating home values based on recent sales data and related home and community characteristics.

Classification determines which category an object should be assigned to based on a training dataset. ArcGIS includes many classification methods for use on remotely sensed data. The tools that use these methods analyze pixel values and configurations to solve problems delineating

↓ Based on the analysis of seven years of traffic accident data, the model predicted areas with the highest risk for accidents. These are shown in red. The analysis considered many factors associated with accidents: weather, time of day, speed limit, proximity to an intersection, and road characteristics. The locations of actual accidents are shown as red/yellow points.

Where loan grades don't effectively explain interest rates

↑ In his Esri Story Maps app, *Mapping the Geography of Online Lending*, author Jonathan Blum used the geographically weighted regression (GWR) tools in ArcGIS to explore the effect of loan grade rankings on average interest rates.

land-use types or identifying areas of forest loss. Maximum Likelihood Classification, Random Trees, and Support Vector Machine are examples of these tools.

Clustering groups observations based on similarities in value or location. ArcGIS includes a broad range of algorithms that find clusters based on one or many attributes, location, or a combination of both attributes and location. These clustering methods can be used for tasks such as segmenting school districts based on socioeconomic and demographic characteristics. Examples of clustering tools in ArcGIS include Spatially Constrained Multivariate Clustering, Multivariate Clustering, Density-Based Clustering, Image Segmentation, Hot Spot Analysis, Cluster and Outlier Analysis tools, and the Space Time Pattern Mining tools.

In addition to ML methods and techniques in ArcGIS tools, ML is used throughout the ArcGIS platform for enabling smart, data-driven defaults, automating workflows, and optimizing results.

For instance, the EBK Regression Prediction method uses principal component analysis (PCA) as a means of dimension reduction to improve predictions. The ordering points to identify the clustering structure (OPTICS) method in density-based clustering tools uses ML techniques to choose a cluster tolerance based on a given reachability plot. The Spatially Constrained Multivariate Clustering tool

↓ The relationship between artificial intelligence, machine learning, and deep learning
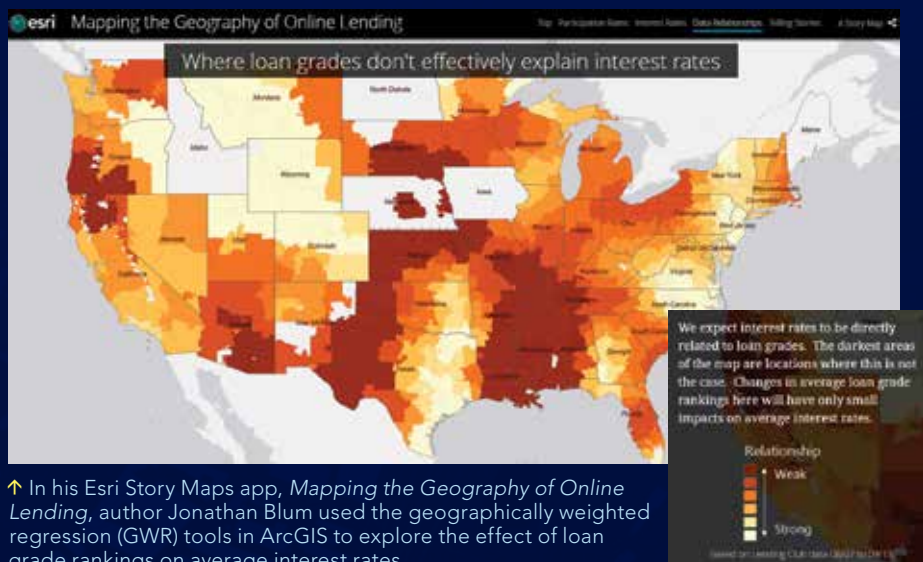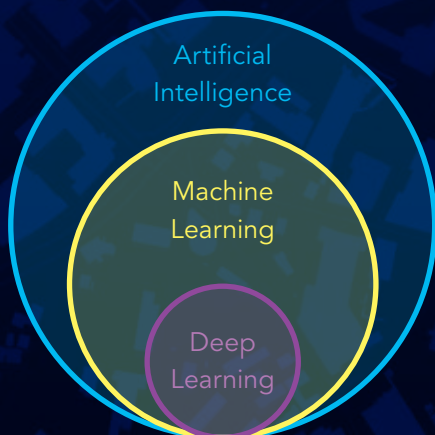
uses an approach called evidence accumulation to provide the user with probabilities related to clustering results.

The field of ML is broad, deep, and constantly evolving. ArcGIS is an open, interoperable platform that allows the integration of complementary methods and techniques in several ways: through the ArcGIS API for Python, the ArcPy site package for Python, and the R-ArcGIS Bridge. This integration empowers ArcGIS users to solve complex problems by combining powerful built-in tools with any ML package they need—from scikit-learn and TensorFlow in Python to caret in R to IBM Watson and Microsoft AI—and still benefit from spatial validation, geoenrichment, and visualization of results in ArcGIS. The combination of these complementary packages and technologies with the systems of record, insight, and engagement that the ArcGIS platform provides is greater than the sum of its parts.

There are many key Esri initiatives for advancing and integrating ML methods across the platform. This road map includes methods such as random forests, neural networks, logistic regression, and time-series forecasting as well as simplified user experiences for integrating with popular ML libraries and packages. A continued focus on distributed processing will play a major role in these advancements.

In addition to building traditional ML into ArcGIS and improving the ease of integrating ML with ArcGIS, Esri is actively

working to broaden the intersection of GIS and ML. This focus on innovation in spatial ML to develop algorithms and approaches that incorporate space into computation will continue to empower ArcGIS users to take advantage of the latest advances in technology and computing while still focusing on solving problems in a fundamentally spatial way.

Explore the many spatial statistics tools that employ machine learning at esriurl. com/spatialstats and visit the Spatial Statistics Forum on GeoNet.

About the Author

Lauren Bennett leads the spatial statistics software development team at Esri. Her role includes providing vision and direction for future analytical tools in the areas of spatial statistics, machine learning, and spatiotemporal analysis as well as involvement in the research, design, validation, documentation, and creation of educational resources for new tools. Other responsibilities include contributing vision for innovative analytical software development across the ArcGIS platform, teaching graduate-level seminars in spatial statistics, and evangelizing the value of spatial analytics. Bennett received a bachelor's degree in geography from McGill University, a master's degree in geographic and cartographic science from George Mason University, and a doctorate in information systems and technology from Claremont Graduate University.