Improving Interpretability in Machine Learning Using Confidence Intervals in ALE Plots

John R. Stevens[1] and Matthew Lister[1,2]

[1]*Department of Mathematics and Statistics, Utah State University*

[2]*Space Dynamics Laboratory, Utah State University*

ABSTRACT

Machine learning models that predict a response based on a collection of predictor variables usually do not provide simple numeric summaries of predictor effects and so are often referred to as black box models. Accumulated local effects (ALE) plots have been developed to allow visual interpretability of predictor effects in such black box models. We present R package AleCI, which improves the original ALE implementation by adding a bootstrapped confidence interval around each prediction showing the range where the true value of the predictor's effect should exist, for categorical as well as continuous predictors. AleCI is applicable across a variety of machine learning models and updates the graphing capabilities of the original implementation by using ggplot2.

Keywords: machine learning, statistical visualization, model interpretability

Guided Data Visualization via Random Forests and Manifold Learning

Kevin Moon
Utah State University

ABSTRACT

The manifold assumption has been used in many machine learning applications to combat the curse of dimensionality and to visualize the data structure. Most manifold learning methods are unsupervised and typically focus on preserving the dominant structure and variation in the data. In many cases, we wish to analyze the data in a supervised setting with respect to expert-provided data labels. Most supervised manifold learning methods exaggerate the separation between data points of different classes, distorting the true structure of the data. In this talk, I will present RF-PHATE, a supervised dimensionality reduction method that preserves the true structure of the variables that are relevant for the supervised task.

Keywords: data visualization; manifold learning; dimensionality reduction

Back to Sessions List

AI-Assisted Data Visualization and Analytics

Kwan-Liu Ma

Computer Science, University of California at Davis

ABSTRACT

Today AI technology is transforming every field, including the domain of visualization. As a powerful tool for data-driven problem-solving, decision-making, and storytelling, visualization must be thoughtfully designed to ensure both effectiveness and performance. I will show how AI and machine learning can help enhance the visualization and analytical reasoning process, drawing from projects conducted by my research group. These technologies hold significant promise in facilitating critical steps of data analysis, guiding in the visualization design space, and optimizing visual outputs to improve user experience and performance.

Keywords: Machine learning; artificial intelligence; data analytics; visualization

Back to Sessions List

iISOMAP: Nonlinear Dimensionality Reduction and Visualization for Interval-Valued Data via Geodesic Distance Preservation

Han-Ming Wu*

Department of Statistics, National Chengchi University

ABSTRACT

Dimensionality reduction for interval-valued data remains an active area of research within symbolic data analysis (SDA). While most existing approaches have primarily focused on adapting linear methods, such as principal component analysis (PCA), to interval data, this study introduces a novel nonlinear approach, namely interval ISOMAP (iISOMAP), which extends the isometric feature mapping (ISOMAP) technique to interval-valued data, explicitly focusing on geodesic distance preservation to uncover the intrinsic geometric structure of datasets. Unlike PCA, ISOMAP is a nonlinear dimensionality reduction (NLDR) method that reconstructs manifold structures by leveraging shortest-path distances on a neighborhood graph. The core innovation of iISOMAP lies in its use of interval multidimensional scaling (MDS) to accurately estimate and preserve geodesic distances between interval-valued data points. For visualization, the maximum covering area rectangle (MCAR) method is employed, projecting interval objects onto a two-dimensional NLDR subspace while maintaining their inherent uncertainty. We evaluate the performance of iISOMAP on both simulated and real-world datasets, comparing it against interval PCA and interval MDS to demonstrate its superior ability to capture nonlinear manifolds. Furthermore, we propose a novel aggregation method that summarizes a dataset's nonlinear structure into interval-valued representations, enhancing iISOMAP's applicability across a range of symbolic data analysis tasks.

Keywords: Interval multidimensional scaling; Isometric feature mapping; Symbolic data analysis

Back to Sessions List