

R12b Kernel PCA × Kernel SHAP による特徴抽出手法の新展開

大久保宏真 (筑波大学), 竹内努 (名古屋大学, 統数研), 赤穂昭太郎 (産総研, 統数研), 五十嵐康彦 (筑波大学), 久野成夫 (筑波大学), 斉藤俊貴 (静岡大学), 中島拓 (名古屋大学), 原田ななせ (国立天文台)

主成分分析 (Classical PCA) は天文学において頻繁に使用される教師なし機械学習手法である。Classical PCA とは、データの分散が大きい方向に新たに軸を作成し (PC 軸)、その軸でデータを評価し直す次元削減・特徴抽出手法である。銀河の画像データに対して Classical PCA を適用するときは、特徴抽出による物理状態の推定をメインとする場合がほとんどである。この場合、分散の大きい PC 軸で抽出された情報は、すでに物理的な研究から知られている場合が多い。そのため、未知の現象を調べたい場合には、分散の小さい PC 軸を確認する必要がある。しかし、Classical PCA は線形的な特徴抽出手法であるため、銀河の画像データのような複雑な構造を持つデータに対しては、正確に特徴を抽出できない場合がある (e.g., Okubo et al. in prep)。

そこで、我々は非線形的特徴抽出手法のカーネル主成分分析 (Kernel PCA) が、Classical PCA よりも詳細に特徴抽出できるのかを検証した。本研究では既に Classical PCA による研究が実施されている NGC 1068 に対して Kernel PCA を適用した。これにより、Classical PCA と Kernel PCA によって抽出された特徴の比較を行う。その結果、Classical PCA では AGN アウトフローとバーエンドの情報が混合し、一つの特徴として抽出された。一方で Kernel PCA では、それぞれの情報を分離して抽出することに成功した。また、ブラックボックスなモデルに説明可能性を与える eXplainable AI (XAI) の Kernel SHAP を適用した結果、AGN アウトフローの影響により CN, CCH, HNC, HCN, HCO⁺, CS が enhance している可能性をデータ科学的に示唆した。以上の結果から、Kernel PCA と Kernel SHAP の組み合わせは、Classical PCA よりも適切に特徴を分離する手法であると言える。