Contribution ID: **39**   Type: **not specified**

Feature selection with Distance Correlation

Friday 4 November 2022 10:20 (20 minutes)

Feature selection algorithms can be an important tool for AI explainability. If the performance of neural networks trained on low-level data can be reproduced by a small set of high-level features, we can hope to understand "what the machine learned". We present a new algorithm that selects features by ranking their Distance Correlation (DisCo) values with truth labels. We apply this algorithm to the classification of boosted top quarks and use a set of 7,000 Energy Flow Polynomials (EFPs) as our feature space. We show that our method is able to select a small set of high-level features, with a classification performance comparable to the state-of-the-art top taggers.

Primary authors: SHIH, David; KASIECZKA, Gregor (Hamburg University (DE)); DAS, RANIT

Presenter: DAS, RANIT

Session Classification: Interpretability