OneModel Academy

One AI

Dimensionality Reduction

Machine Learning

Hayley Bresina
One AI Client Enablement

08 Jul 2024

Topics Covered

- Overview of dimensionality & dimensionality reduction
- How dimensionality impacts machine learning models
- Available configuration options for dimensionality reduction in One AI
- Default configuration for dimensionality reduction in One AI
- How to configure dimensionality reduction in One AI

Learning Outcomes

You will:

- Understand what dimensionality is & why reducing dimensionality is necessary for improving model performance
- Know the default configuration for dimensionality reduction in One AI & how to configure & customize methods and feature numbers if desired
- Distinguish between various dimensionality reduction techniques & understand how to apply them effectively to enhance model performance

Dimensionality & Dimensionality Reduction

Dimensionality & Dimensionality Reduction

- **Dimensionality**: number of input variables or features in a dataset used for training a ML model
  - **High-dimensional datasets** contain a large number of features; often including several rows of null or constant data
  - High-dimensional data negatively impacts computational efficiency & performance
- Dimensionality reduction addresses issues with high dimensional data
  - Reduces the number of input features to simplify the dataset while retaining as much relevant information as possible
  - One AI model datasets typically contain 50-500 features; 5-15 is optimal

Dimensionality Reduction

- One AI optimized dimensionality through **filter & wrapper methods**
  - **Filter method -** evaluates each feature individually using statistical tests to see how relevant it is to the target variable
    - Ranks features based on individual relevance & sends the best features to the next step
  - **Wrapper method -** a predictive model tests different combinations of these important features to find the best set by trying all of the features, & then removing 1 at a time to see if the model performs as well without it. If it does, that feature is dropped.
    - More computationally expensive than the filter method

Dimensionality Impacts on Models

Dimensionality Impacts on Models

- Less is more - there's an ideal number of input variables that improves model performance; so we need dimensionality reduction
    - Avoid the **curse of dimensionality**, where with more features, the dataset becomes more complex, requiring more data points to explore & represent all possible feature combinations & capture the dataset's variability
    - Prevent sparsity & overfitting
    - Visualize the data more easily, helping end users interpret complex patterns more effectively
    - Allows use of larger datasets without excessive runtime or computational demands

Configuration Options in One AI

Configuration Options In One AI

- Configuration provides control over how many features are tried & selected
- Options
  - Disable dimensionality reduction
  - Configure **filter method**: Mutual info, Chi-square test, F-test (ANOVA)
  - Configure filter number of features - **maximum** number of features that will be selected in the model
  - Wrapper method can’t be configured; **recursive feature elimination** is the only option
  - Configure wrapper minimum number of features - **minimum** number of features that will be selected in the model

Default Configuration in One AI

Default Configuration In One AI

- The default configuration in One AI generally performs well & is based on best practices
  - Default filter method: **mutual_info**
  - Filter number of features: **5, 10, & 15** are tried & the best results is selected
  - Default wrapper method: **recursive feature elimination (rfe)**
  - Wrapper min features: **5**

Configuring in One AI

Thanks for watching!