Feature Selection and Reduction techniques in machine learning

Course: Machine Learning
Section:
Faculty: MD. SAEF ULLAH MIAH
Presenter: MD. Farhan Sadik
ID: 21-44403-1

INDEX

DEFINITIONS

Feature Selection is the process of selecting a subset of the most relevant characteristics (variables or attributes) from a wider range of possible features that aims to decrease the likelihood of overfitting, increase interpretability of the models, and boost machine learning models' performance and efficiency.

Feature reduction which is often referred to as dimensionality reduction, is the process of lowering the quantity of input variables or features in a dataset while maintaining the majority of data that is relevant. In machine learning and data analysis in particular, it is a commonly utilized method for feature engineering and data preprocessing.

DIFFERENCES BETWEEN FEATURE SELECTION AND REDUCTION

Feature Selection	Feature Reduction
• Choosing subset of relevant features	• Transformation the dimension of the dataset to a lower order
• Gives a smaller set of features while discarding the rest	• Combines features into a smaller set of features
• Utilizes filer methods, wrapper methods & embedded methods	• Utilizes techniques such as Principal Component Analysis, Linear Discriminant Analysis, Autoencoders etc.

FEATURE SELECTION USAGE

- Maintaining Transparency/Interpretability which allows retaining of the original features which are deemed relevant, indicating which variables affect the model's predictions
- Domain knowledge and good understanding of the relevant features are known
- Feature number is greater (reducing the complexity of the model)
- Improving Model Training time
- Establishing a clear understanding of the relationship between features and target variable

FEATURE REDUCTION USAGE

- Handling dataset with high dimensionality
- Highly correlated features are present in the dataset: capturing the underlying structure of the data by creating new features which are uncorrelated
- Visualizing high dimensional data in lower dimensions
- Feature importance is uncertain as in not knowing which features are most relevant
- Noise reduction by focusing more on the relevant information in the fewer and more meaningful dimensions.

FEATURE SELECTION TECHNIQUES

1. Correlation-based
2. Recursive feature elimination
3. L1 regularization

FEATURE REDUCTION TECHNIQUES

1. Principal component analysis
2. Linear discriminant analysis
3. t-Distributed Stochastic Neighbor Embedding
4. Autoencoders

IMPLEMENTATION OF FEATURE SELECTION AND REDUCTION

A sample of this dataset , which contained data related heart attack risk factors, was used during both feature selection and reduction

The implementation of the feature selection and reduction techniques are here

For feature selection, RFE(Recursive Feature Elimination) was used as the dataset contained both numerical and non-numerical data on a random forest .

For feature reduction, PCA (Principal Component Analysis) was used since the dimensionality of the data was considerably high.

Usage of RFE yielded a model with 62-68% accuracy and usage of PCA yielded a model with 64-69% accuracy.

A SHORT DESCRIPTION OF RFE

It is a feature selection technique commonly used in machine learning. It works by iteratively fitting a model to the data, ranking the importance of each feature, and eliminating the least important feature in each iteration until the desired number of features is reached. RFE is often used in combination with models that assign feature importance scores, such as decision trees or support vector machines (SVMs).

A SHORT DESCRIPTION OF PCA

Principal Component Analysis (PCA) is a dimensionality reduction technique commonly used in machine learning and data analysis. Its primary goal is to reduce the dimensionality of a dataset while preserving as much of the data's variability as possible.

The steps to implement PCA is as follows:

1. Standardization of the data: Since PCA implies that the data is centered at the origin, standardizing the data—that is, dividing by the standard deviation and subtracting the mean—will ensure that each feature has a standard deviation of 1 and a mean of 0.
2. Computing the covariance matrix: The covariance matrix measures how different features vary together. It's computed from the standardized data. For a dataset with n samples and m features, the covariance matrix Σ is an m × m matrix where Σ_ij is the covariance between features i and j.

$$\frac{1}{n-1}\sum_{i=1}^{n}(Xi - X')(Xi - X')^T$$

Here, X_i is the standardized data point I and X' is the mean vector of the data

A SHORT DESCRIPTION OF PCA (CONTD.)

1. **Computing eigenvectors and eigenvalues:** The primary components of the data are represented by the eigenvectors, and their significance is shown by the eigenvalues. These can be calculated with specialist eigenvalue solvers or through Singular Value Decomposition (SVD). The eigenvectors are normally arranged by decreasing eigenvalues and create an orthonormal basis. The eigenvector with the biggest eigenvalue is represented by the first principal component, the second by the second largest, and so on.
2. **Choosing the number of principal components:** The number of principal components is specified before variance is examined. Examining the explained variance, or the percentage of the total variance that each principal component explains, is a popular strategy. Either a threshold percentage value is chosen or the number of principal components is specified as necessary.
3. **Projecting the data:** The final step is to project the data onto the selected principal components. This is done by taking the dot product of the data with the eigenvectors of the selected principal components. $X_{new} = X.V$. Here, X_{new} is the reduced-dimension data, X is the standardized data, V is a matrix where each column is a selected eigenvector

BENEFITS OF FEATURE REDUCTION

- Feature reduction simplifies the dataset by reducing the number of features. This has several advantages such as computational and memory efficiency
- Improved model performance as feature reduction can mitigate this risk by focusing on the most informative features, making models more generalizable and simpler for interpretation
- Noise reduction as the impact of noisy features is eliminated or minimalized
- Machine learning models can be trained faster when working with a reduced set of features

POTENTIAL ISSUES OF NOT IMPLEMENTING FEATURE REDUCTION

- A model may learn noise and yield less accurate predictions if the dimensionality is not reduced and if it contains characteristics that are unnecessary or redundant can lead to overfitting
- Without feature reduction, high-dimensional datasets will require more time and computational resources to train and deploy machine learning models
- Irrelevant or redundant features in the dataset can obscure meaningful patterns, making it harder to extract valuable insights from the data
- Processing high-dimensional data may not be feasible due to memory and processing constraints, especially in resource-constrained environments

THANK YOU