FOUNDATIONS OF MACHINE LEARNING
FEATURE SELECTION

Lecturer: Moustapha CISSE

African Masters of Machine Intelligence

Group Members:
Abigail Naa Amankwaa Abeo
Etienne Ntumba
Jean Robin Raherisambatra
Souleymane Balde

AIMS African Institute for Mathematical Sciences SENEGAL

Outline

What is Feature Selection

- Feature Selection involves selecting a subset of relevant features from a larger set of available features.
- The process of reducing the number of input variables when developing predictive models.

AIMS African Institute for Mathematical Sciences SENEGAL

Example of feature selection

Feature Selection

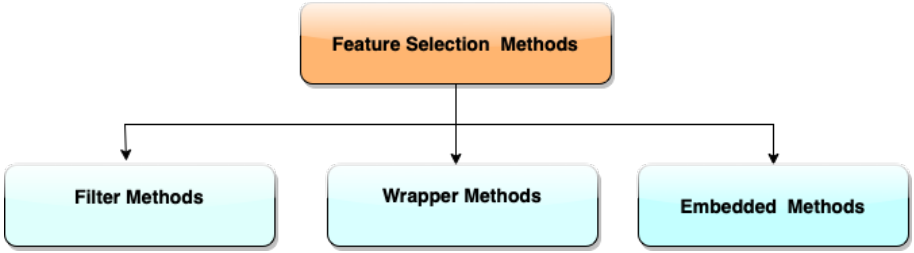
Name	Math	Chemistry	Maths	Physics	General Test	Result
A	70	60	70	50	70	Pass
B	60	80	60	50	70	Pass
C	40	65	40	50	60	Fail
D	80	55	80	50	60	Pass
E	30	60	30	50	80	Fail

Math	Chemistry	General Test	Result
70	60	70	Pass
60	80	70	Pass
40	65	60	Fail
80	55	60	Pass
30	60	80	Fail

Why Feature Selection

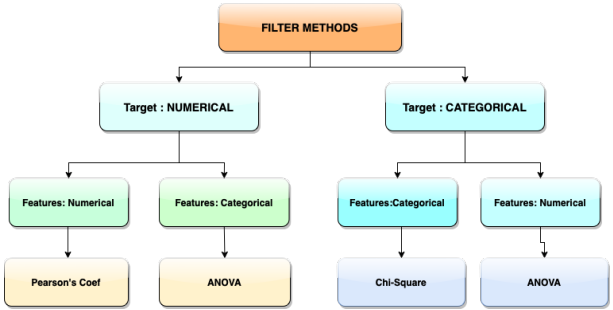
- To remove irrelevant and redundant features
- To avoid overfitting
- To improve model performance
- To have faster training and prediction

DIFFERENT FEATURE SELECTION METHODS



Filter Methods

- Rely on statistical measures to rank the importance of features.
- Select the top features based on ranking.



AIMS African Institute for Mathematical Sciences SENEGAL

Pearson's Correlation

- It is used as a measure for quantifying linear dependence between two continuous variables X and Y.
- Its value varies from -1 to $+1$

$$\rho_{X,Y} = \frac{cov(X, Y)}{\delta_X \delta_Y} \tag{1}$$

AIMS African Institute for Mathematical Sciences SENEGAL

ANOVA

- ANOVA can be used as a filter method in feature selection by computing the F-statistic for each feature, which measures the difference in mean values across different groups of samples.

- ANOVA only captures linear relationships between features and the target variable, and may not work well if there are non-linear relationships present.

Chi-Squared

- It is used as a measure for quantifying linear dependence between two categorical variables X and Y.
- Assess the significant difference based on p-value(< 0.05)

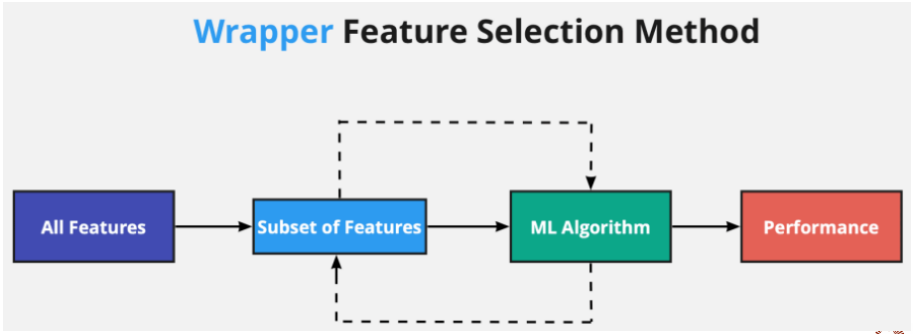
$$X^2 = \frac{\sum (O_i - E_i)^2}{E_i} \tag{2}$$

AIMS African Institute for Mathematical Sciences SENEGAL

Wrapper Methods

- These methods create several models with different subsets of input features.
- The selected features result in the best performing model in accordance with the performance metric.
- These methods follow a greedy search approach by evaluating all the possible combinations of features.

Wrapper Methods



Wrapper Methods

- Types of Wrapper Methods

1. **Forward Feature Selection** : Starts with a (usually empty) set of variables and adds variables to it, until some stopping criterion is met.

2. **Backward Feature Selection** : Starts with a (usually complete) set of variables and then excludes variables from that set, again, until some stopping criterion is met.

L1 REGULARIZATION

- This method consists of adding a penalty term to the cost function of a machine learning algorithm that brings the model to use fewer features in its predictions.
- L1 regularization tends to push the coefficients of some features to zero, effectively removing them from the model.
- L1 regularization will help to have a sparse matrix

$$L1\,regularisation = lossFunction + \lambda \| W \|_1$$

Sparsity

- L1 Regularization forces the coefficient to tend towards zero, this will create a sparse matrix, which contains high number of zeros.

Users					Movies						Target
A	B	C	D	E	Parasite	Joker	Avengers	Spotlight	The Great Beauty	There will be blood	Rating
1	0	0	0	0	1	0	0	0	0	0	5
1	0	0	0	0	0	1	0	0	0	0	4
1	0	0	0	0	0	0	1	0	0	0	4
0	1	0	0	0	1	0	0	0	1	0	2
0	1	0	0	0	0	0	0	1	0	0	4
0	1	0	0	0	0	0	0	0	1	0	3
0	0	1	0	0	0	0	1	0	0	0	5
0	0	0	1	0	0	0	0	0	0	1	4
0	0	0	0	1	0	0	1	0	0	0	4

AIMS African Institute for Mathematical Sciences SENEGAL

Linear Model

- Linear regression models aim to predict the outcome based on a linear combination of the predictor variables given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

- The loss function is obtained by minimizing the squared difference between the actual value and the predicted value of y.

$$minimize\{\sum_{j=1}^{N}(y_j - \beta_0 - \sum_{i=1}^{n}\beta_i x_{ij})^2\}$$

AIMS African Institute for Mathematical Sciences SENEGAL

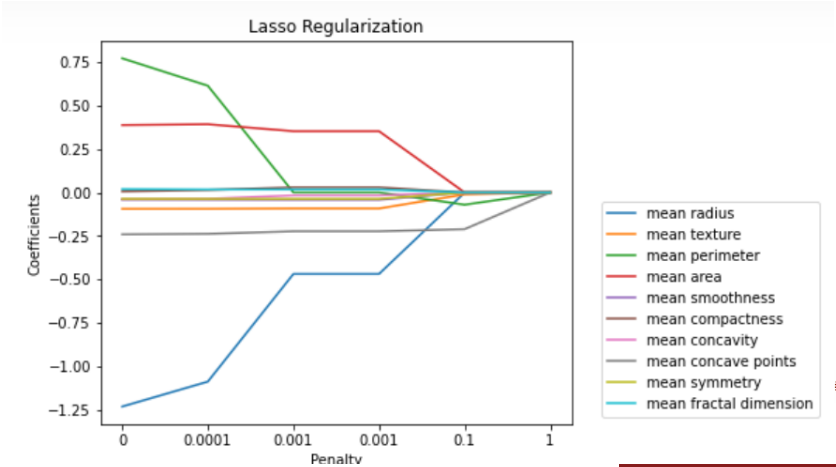
L1 Regularisation

In high-dimensional feature spaces, that is, if the dataset contains many features, linear models are likely to overfit the data. To avoid this, the search for the optimal coefficients is done with regularization. One of the types of regularization is that of the L1 norm, also called lasso regularization in the case of a linear regression.

$$minimize\{\sum_{j=1}^{N}(y_j - \beta_0 - \sum_{i=1}^{n}\beta_i x_{ij})^2 + \lambda \sum_{i=1}^{n}|\beta_i|\}$$

AIMS African Institute for Mathematical Sciences SENEGAL

Observation



Implementation

- Find the different implementations in the Jupyter Notebook

Conclusion

- Feature selection presents high capacity of handling overfitting issues through removal of irrelevant features.
- The above methods are using for supervising learning, however feature selection can also be used in unsupervised learning.

References

Review on wrapper feature selection approaches, El Aboudi, Naoual and Benhlima, Laila, 2016 International Conference on Engineering & MIS (ICEMIS), 1–5,2016

Feature Selection methods in Machine Learning, https://deepblade.com/feature-selection-methods-in-machine-learning/, 2022

Feature Selection Methods and How to Choose Them, https://neptune.ai/blog/feature-selection-methods, 2023

AIMS African Institute for Mathematical Sciences SENEGAL

References

L1 Norm
Regularization and Sparsity Explained for Dummies, https://blog.mlreview.com/l1-norm-regularization-and-sparsity-explained-for-dummies-5b0e4be39

Introduction to Feature Selection methods with an example (or how to select the right variables?), https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to#:~:text=Filter,2016-2023

References

Fighting Overfitting With L1 or L2 Regularization: Which One Is Better?, https://neptune.ai/blog/fighting-overfitting-with-l1-or-l2-regularization,2023

AIMS African Institute for Mathematical Sciences SENEGAL