AFRICA DATA SCHOOL

Feature Selection.

Presented by: JACQUES PATRICKS K.

LAW OF PARSIMONY

The best explanation to a problem is that which involves the **fewest** possible **assumptions**.

AFRICA DATA SCHOOL

What is Feature Selection?

- The process of selecting a subset of relevant features for use in model construction.
- Also known as variable selection, attribute selection or variable subset selection.
- A feature in case of a dataset simply means a column.

AFRICA DATA SCHOOL

Why use Feature Selection?

- It enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the right subset is chosen.
- It reduces overfitting.

AFRICA DATA SCHOOL

Taxonomy of Feature Selection

1. Filter Method
2. Wrapper Method
3. Embedded Method

AFRICA DATA SCHOOL

Filter Method

- Picks up the *intrinsic* properties of the features.
- Measured via *univariate* statistics instead of cross-validation performance.
- When dealing with *high-dimensional* data, it is computationally cheaper to use filter methods.
- Filtering is done using correlation matrix and it is most commonly done using *Pearson correlation.*

AFRICA DATA SCHOOL

Go to Code

Wrapper Method

- Needs one machine learning algorithm and uses its performance as evaluation criteria.
- Feed the features to the selected ML algorithm and based on the model performance you *add/remove* the features.
- An *iterative* and computationally expensive process but it is *more accurate* than the filter method.
- Methods: Backward Elimination, Forward Selection, Bidirectional Elimination and RFE.

AFRICA DATA SCHOOL

Go to Code

i. Backward Elimination

- We feed all the *possible features* to the model at first.
- We check the performance of the model and then iteratively remove the *worst performing features* one by one till the overall performance of the model comes in acceptable range.
- The performance metric used here to evaluate feature performance is *pvalue*.

AFRICA DATA SCHOOL

Go to Code

Recursive Feature Elimination (RFE)

- Uses *accuracy metric* to rank the feature according to their importance.
- RFE method takes the *model to be used* and the number of required features as input.
- Then gives the *ranking* of all the variables, 1 being most important.
- Also gives its *support*, True being relevant feature and False being irrelevant feature.

AFRICA DATA SCHOOL

Go to Code

ii. Embedded Method

- Iterative in a sense that takes care of each iteration of the model training process.
- Carefully extract those features which *contribute the most* to the training for a *particular iteration.*
- Regularization methods penalize a feature given a *coefficient threshold.*
- For irrelevant feature, lasso penalizes it's coefficient and make it 0. features with coefficient = 0 are removed.

AFRICA DATA SCHOOL

Go to Code

Categorical Features

Chi-square Test of Independence

- Calculates Chi-square between each feature and the target. (significant relationship)
- Selects the desired number of features with the best *Chi-square scores*.
- Conditions: *Categorical variable* and values should have expected *frequency > 5*.
- Hypothesis testing test with 2 hypotheses present; the *Null* Hypothesis and the *Alternative* Hypothesis.

AFRICA DATA SCHOOL

Go to Code

CONCLUSION

There are many other methods of feature selection. There are hybrid methods too that use both filtering and wrapping techniques.

AFRICA DATA SCHOOL