Feature Selection and Transformation

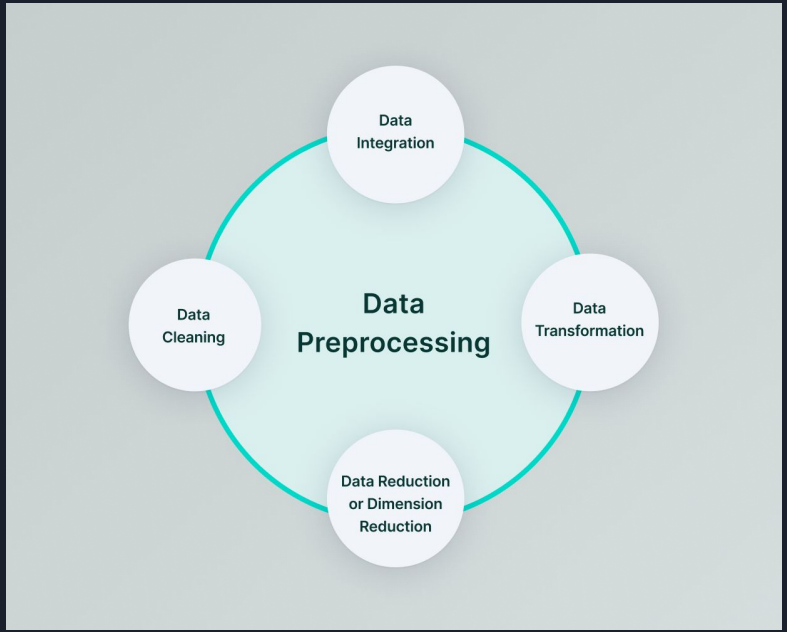
Authors: Arushan and Victoria
Date: November 24, 2021

TABLE OF CONTENT

1.0 Data Preprocessing

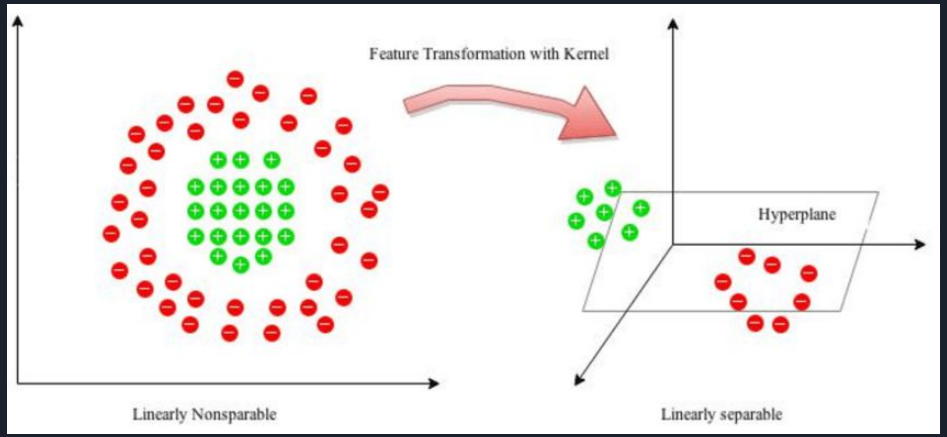
Data preprocessing is an important part of Machine Learning as the quality of data can directly affects the ability of a model to learn

Feature Selection and Transformation are two major parts in Data preprocessing



2.0 What Is Feature Transformation

- “Feature transformation is the process of modifying your data but keeping the information” - Renan Lolico (Towards Data Science)
- Feature transformation allows us to manipulate data to increase a model performance
- There are many Machine Learning techniques and methods to transformation data like Data Normalization, and Data Standardization, and One Hot Encoding



2.1 Data Normalization

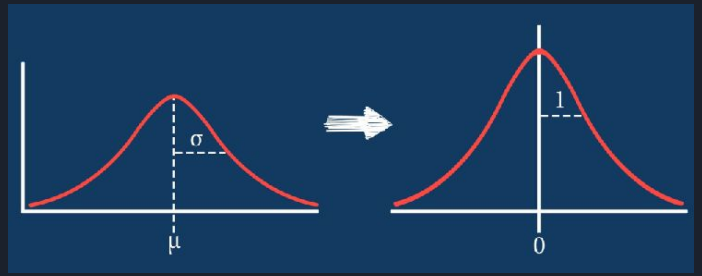
- Data Normalization is the process of converting an actual range of values which a numerical feature can take, into a standard range of values, typically in the interval [0,1]
- Data Normalization typical increase the speed for optimization algorithms, such as gradient descent

$$\bar{x}^{(j)} = \frac{x^{(j)} - min^{(j)}}{max^{(j)} - min^{(j)}}$$

2.2 Data Standardization

- Data Standardization is the a process in which the feature values are rescaled to produce a standard normal distribution with mean of 0 and standard deviation of 1

$$\hat{x}^{(j)} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}}$$



2.3 One Hot Encoding

- One Hot Encoding is a process by which categorical feature are converted into numerical feature
- Increase the dimensionality of your feature vector

id	color
1	red
2	blue
3	green
4	blue

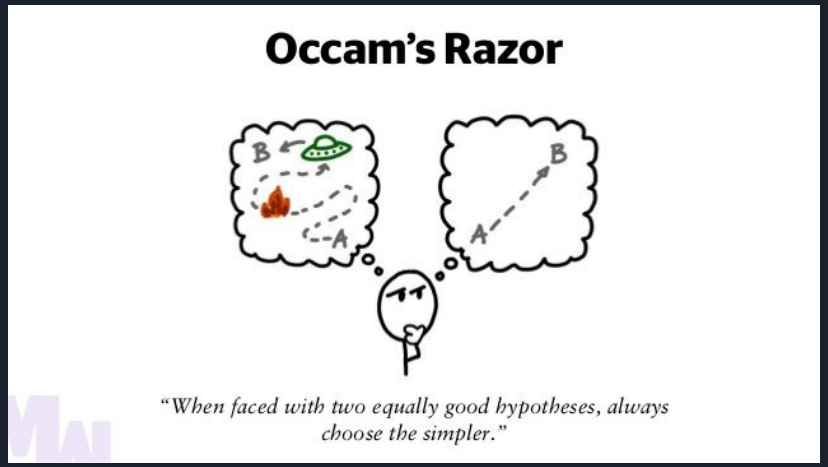
id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

3.0 What Is Feature Selection

The primary purpose of the feature selection to remove redundant and non-informative features.

3.1 The Benefit Of Feature Selection

- Reduce dimensionality of data to improve machine learning performance (preventing overfitting)
- Improves training speed
- The model is more explainable with less features (Occam's Razor)



3.2 Type of Feature Selection

Supervised feature selection techniques can be categorized into methods

- Filter-Based Feature Selection Methods
  - The statistic correlation between input and output features
  - Filter-based feature selections score each feature and select the features with the largest (or smallest) score.
  - Pearson correlation is an example of Filter-Based Feature Selection

- Wrapper Feature Selection Methods
  - Wrapper methods create many models with different subsets of input features and select those features that result in the best performing model according to a performance metric.
  - A model is wrapper with a specific methods and iterative over the features and remove the least importance feature
  - Recursive Feature Elimination is an example of wrapper method

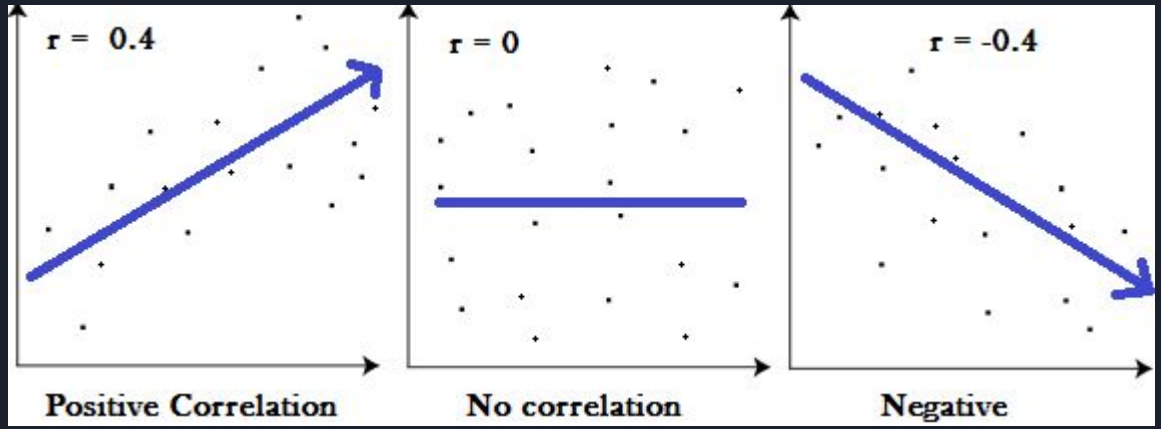
3.3 Pearson’s correlation coefficient

- Pearson's correlation coefficient is also known in statistical models as the r value
- Pearson's correlation coefficient is computed by taking the covariance of two variables and dividing by the product of their standard deviations. The coefficient is not affected by changes of scale in the two variables.
- The value of r can range between -1 to 1

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

3.3 Pearson’s correlation coefficient

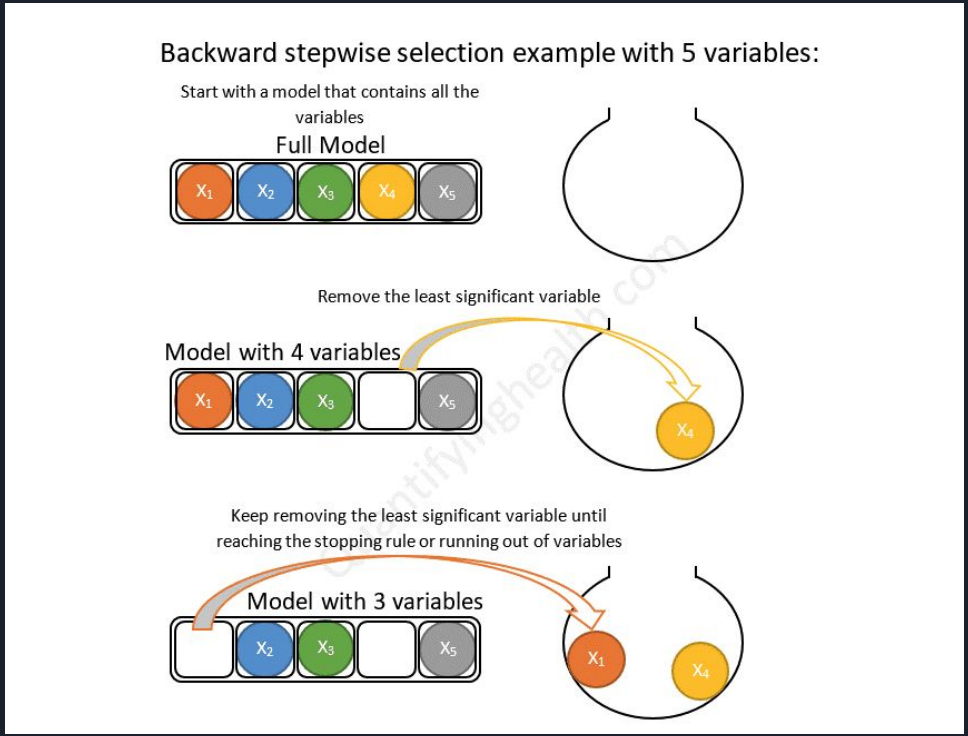
- Positive Correlation (1): both variables change in the same direction.
- Neutral Correlation (0): No relationship in the change of the variables.
- Negative Correlation (-1): variables change in opposite directions.



3.4 Recursive Feature Elimination

- Recursive feature elimination (RFE) is a feature selection method that wrap around fitted model and removes the weakest feature (or features) until the specified number of features is reached
- RFE applies a backward selection process and ranking system to find the optimal combination of features
- The ranking system is based on a feature importance which is determined by coefficient
- RFE works well on small data set.

3.4 Backward Selection



THANK YOU!

If you have questions, we are holding office hours next week at the same time (December 1, 7:30PM - 8:30 PM).

Check out our upcoming Kaggle Competition CAIS X (https://carletonai.com/x/)!

kaggle x CAIS

3.3 Regularization

- Lasso (L1) Regularization is a technique apply to machine learning model to reduce the variance of model.

$$\min_{\mathbf{w},b} \left[C|\mathbf{w}| + \frac{1}{N} \sum_{i=1}^{N} (f_{\mathbf{w},b}(\mathbf{x}_i) - y_i)^2 \right]$$

- Lasso Regularization will remove non-informative feature during the training of model