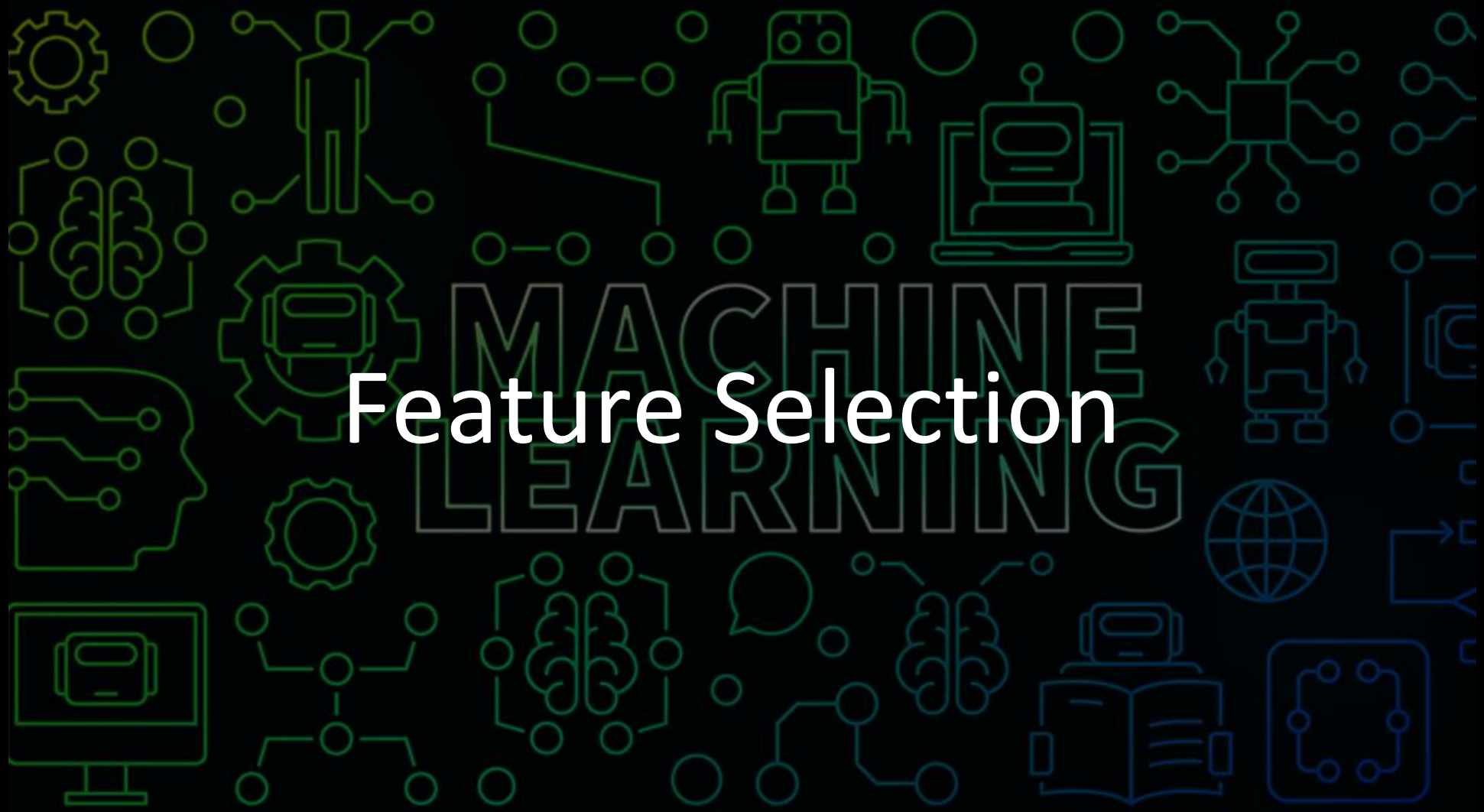
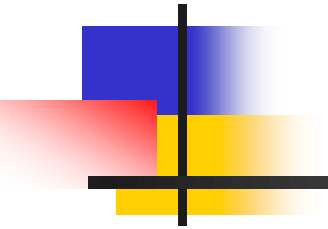


Feature Selection

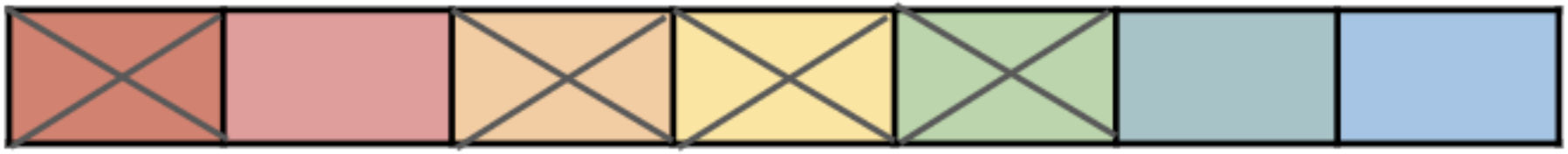


Feature Selection

Feature Selection

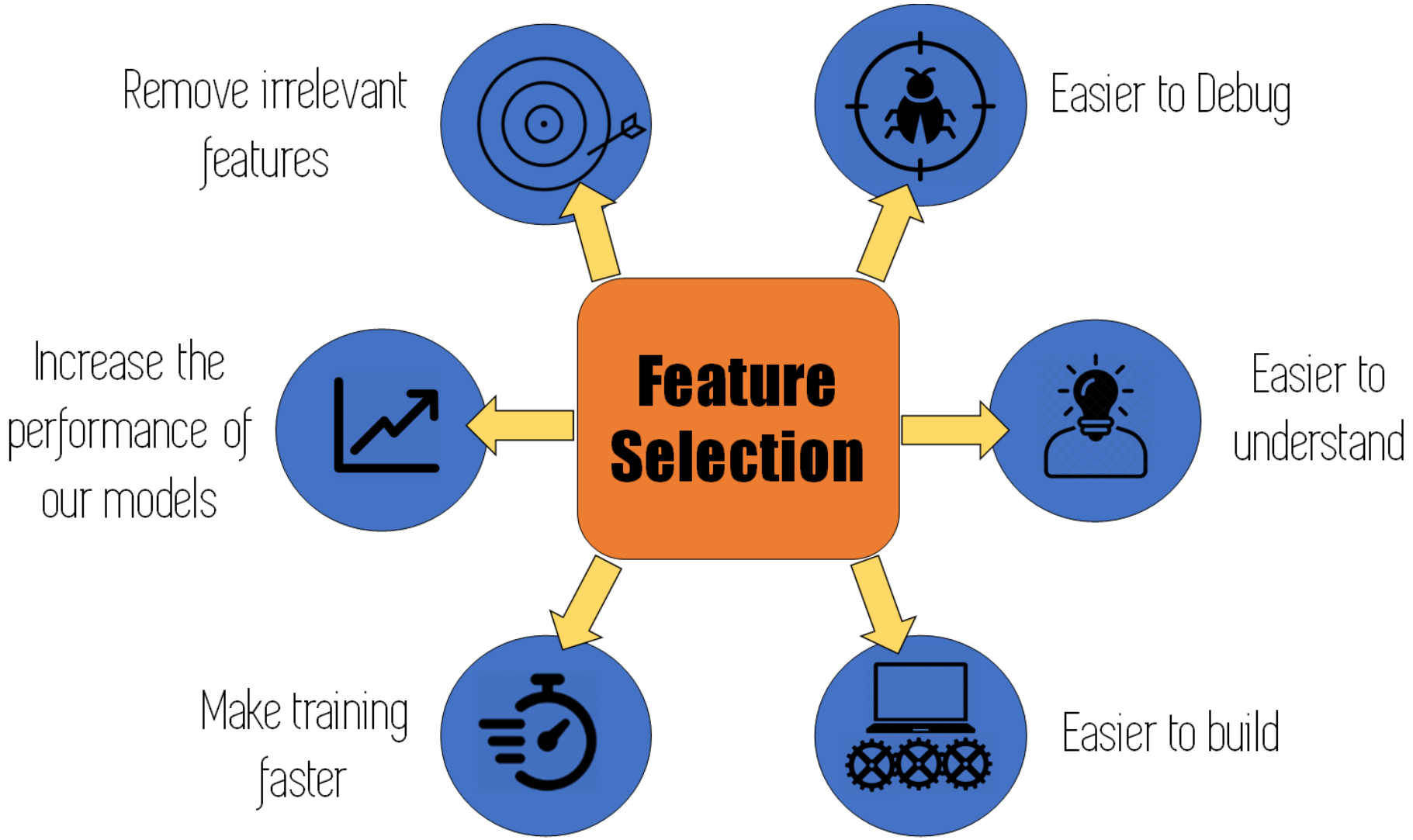
All Features

Feature Selection



Final Features

Why Feature Selection



Feature Selection Types

- Filter methods
- Wrapper methods
- Embedded methods

Filter Methods

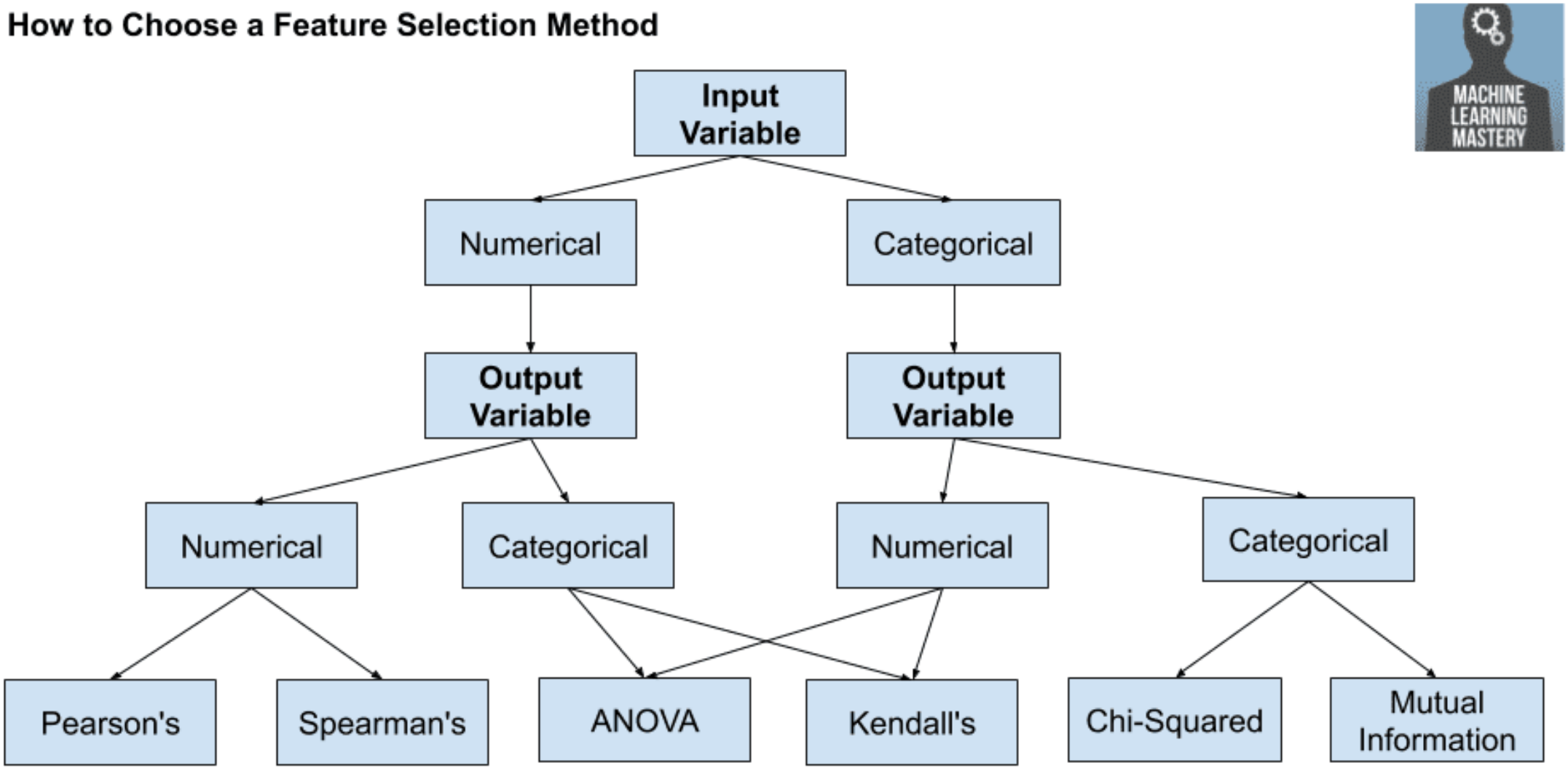
- Filter methods are generally used as a preprocessing step.
- The selection of features is independent of any machine learning algorithms.
- Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable.

Filter Methods

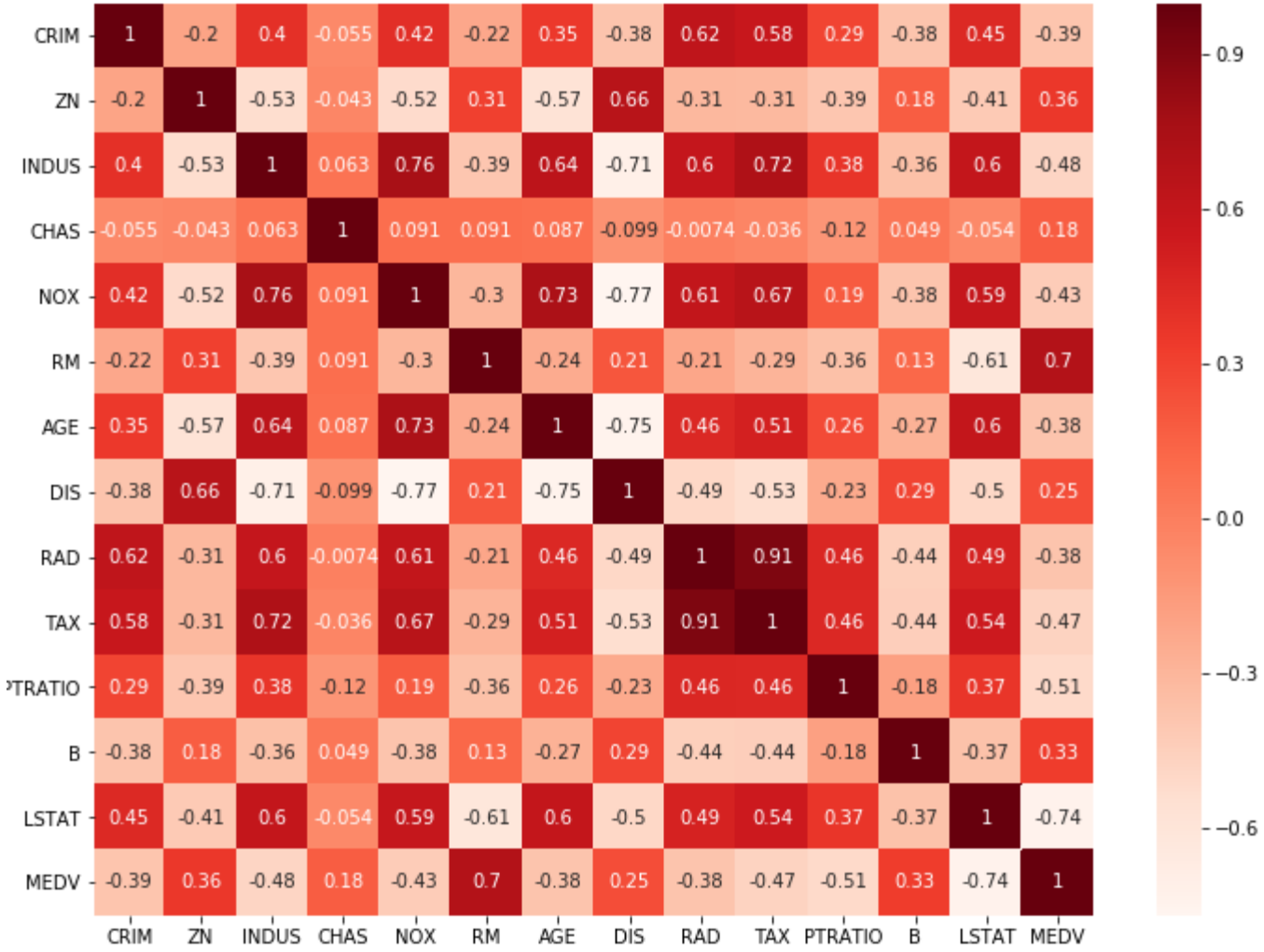
- The correlation is a subjective term here. For basic guidance, you can refer to the following table for defining correlation co-efficients.

Feature\Response	Continuous	Categorical
Continuous	Pearson's Correlation	LDA
Categorical	Anova	Chi-Square

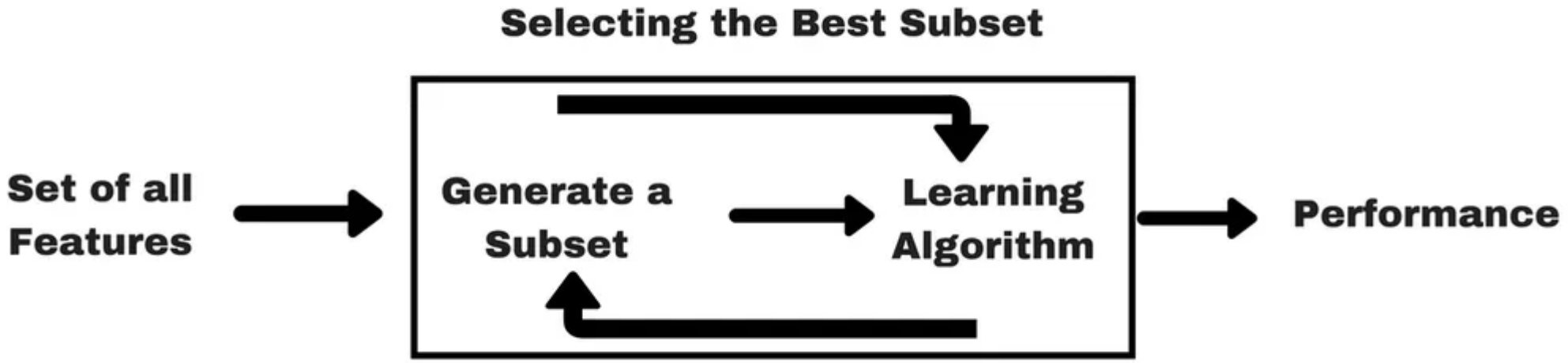
Filter Methods



Filter Methods



Wrapper Methods



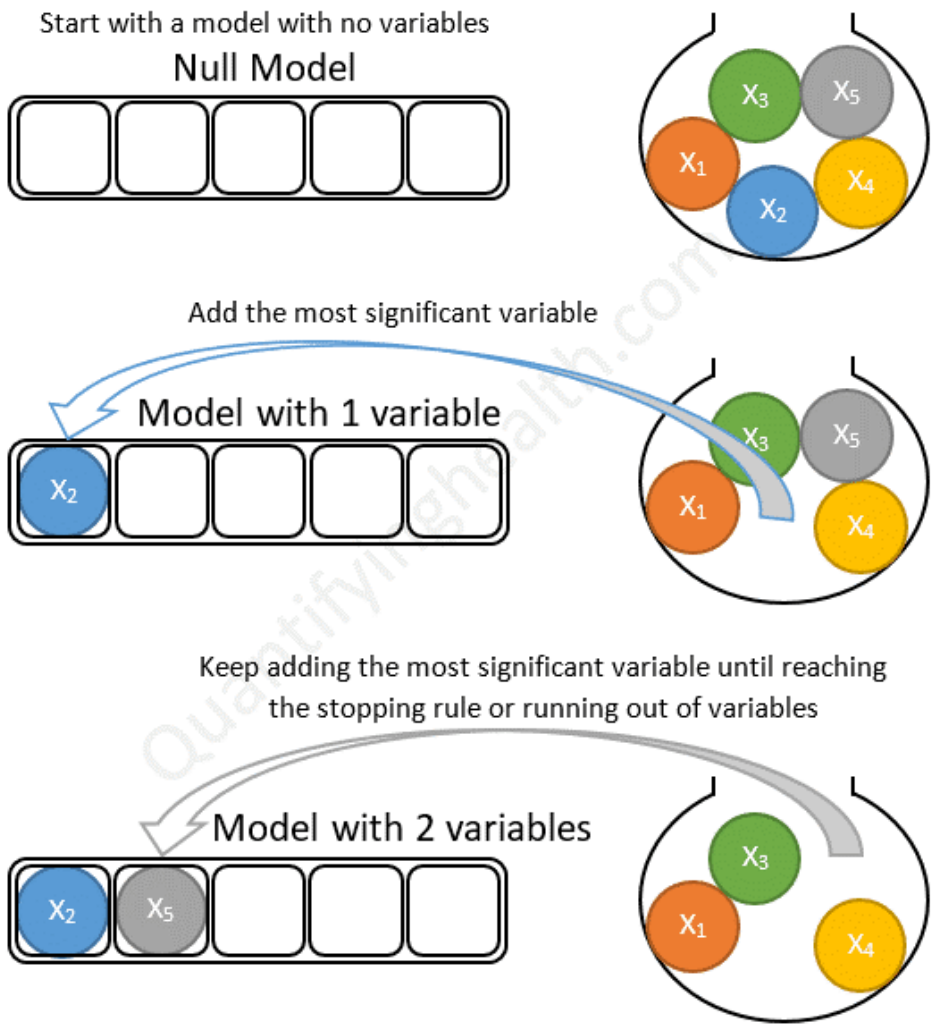
- In wrapper methods, we try to use a subset of features and train a model using them. Based on the inferences that we draw from the previous model, we decide to add or remove features from your subset. The problem is essentially reduced to a search problem. These methods are usually computationally very expensive.

Wrapper Methods

- Some common examples of wrapper methods are forward feature selection, backward feature elimination, recursive feature elimination, etc.
- **Forward Selection:** Forward selection is an iterative method in which we start with having no feature in the model. In each iteration, we keep adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model.
- **Backward Elimination:** In backward elimination, we start with all the features and removes the least significant feature at each iteration which improves the performance of the model. We repeat this until no improvement is observed on removal of features.
- **Recursive Feature elimination:** It is a greedy optimization algorithm which aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination.

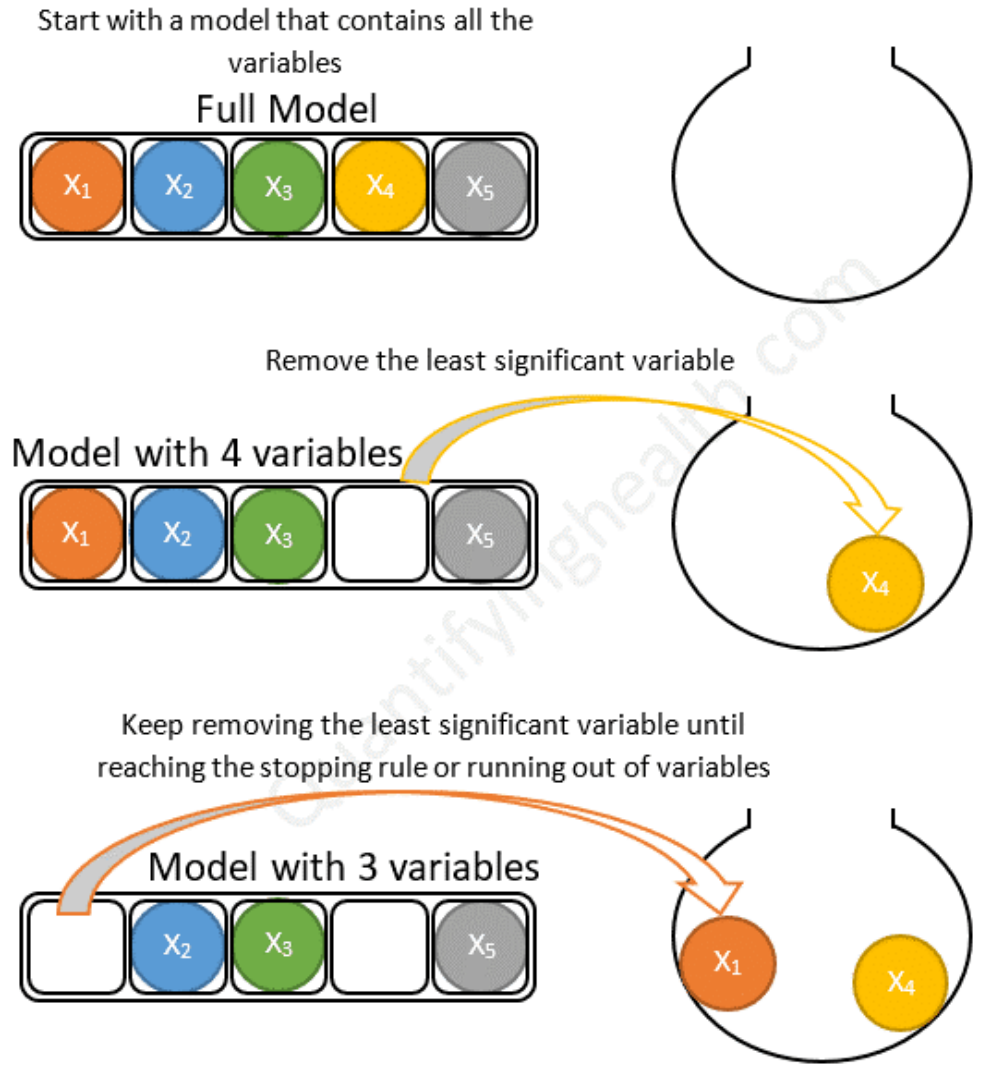
Forward Selection

Forward stepwise selection example with 5 variables:



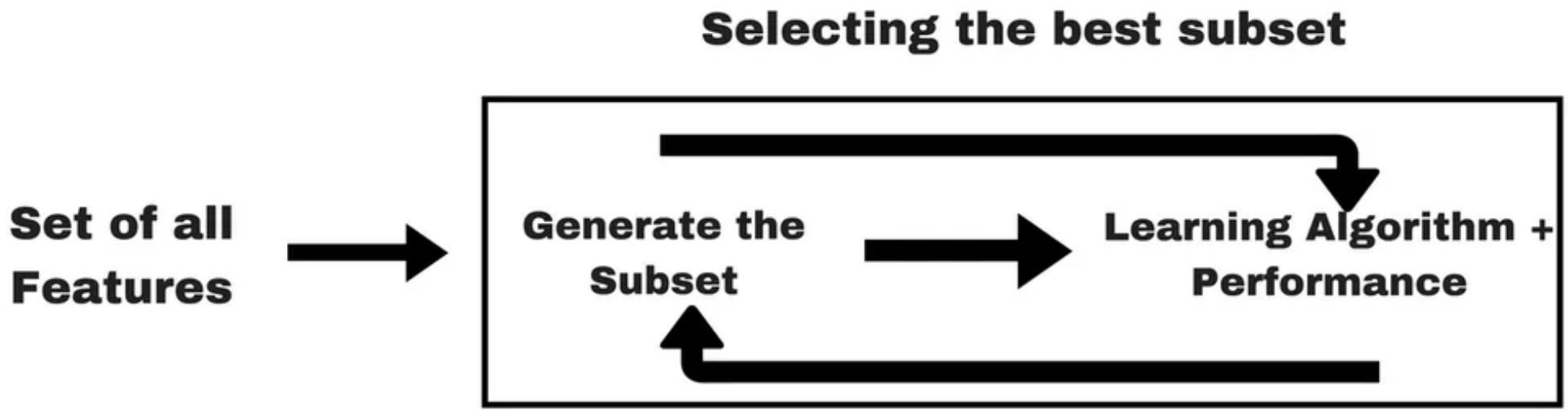
Backward Elimination

Backward stepwise selection example with 5 variables:



Recursive Feature Elimination

Embedded Methods



- Embedded methods combine the qualities' of filter and wrapper methods. It's implemented by algorithms that have their own built-in feature selection methods.

Embedded Methods

- Some of the most popular examples of these methods are LASSO and RIDGE regression which have inbuilt penalization functions to reduce overfitting.
- Lasso regression performs L1 regularization which adds penalty equivalent to absolute value of the magnitude of coefficients.
- Ridge regression performs L2 regularization which adds penalty equivalent to square of the magnitude of coefficients.

Embedded Methods

$$L(x, y) = \sum_{i=1}^{n} (y_i - h_\theta(x_i))^2$$

where $h_\theta x_i = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + \theta_3 x_3^3 + \theta_4 x_4^4$

$$L(x, y) \equiv \sum_{i=1}^{n} (y_i - h_\theta(x_i))^2 + \lambda \sum_{i=1}^{n} \theta_i^2$$

Feature Selection Discussion

- What models need FS?
- What data need FS?
- What happens if you remove most of features?
- Is it really mandatory to perform FS?
- Do we perform FS for unsupervised learning?