In Machine Learning, we use certain techniques to select the best possible (and most relevant) subset of features from our dataset so that our ML models have the best time predicting and analyzing data.

These feature selection techniques allow us:

1. To reduce the dimensionality of the dataset. Too many features can make our models sluggish, and by picking the right features, we can speed up the learning process.
2. To remove features that can confuse our ML model with their noise and irrelevance. This way, our models can focus on what really matters and give us accurate predictions.
3. To understand the ML models' output better, understanding the "why" behind the predictions and gaining valuable knowledge.

Now, how do we go about selecting these features?

1. In "Filter Methods", we rank features based on statistical measures. It's like giving each feature a score to determine how important they are. Simply put, we look at each attribute and determine how strongly correlated it is with the target variable.

2. "Wrapper Methods" are like a game of trial and error. We let our models play around with different subsets of features to find the best combination. We create models for a subset of features and calculate the accuracy to greedily determine which feature combination is the best.

3. With "Embedded Methods", we let our models themselves decide which features are the most useful during the learning process, thus having its own built-in feature selection methods. Examples include Regularization techniques to penalize certain features, and decision trees.

Let's say we have a dataset with various features such as age, income, education level, and gender, and we want to predict whether a person is likely to purchase a particular product.

1. Filter Method

- First, we define a criterion to measure the relevance of each feature. In this example, let's consider the correlation between each feature and the target variable.
- Next, we calculate the correlation between each feature and the target variable. We can use a correlation coefficient like Pearson's correlation or Spearman's rank correlation to quantify the relationship.
- After calculating the correlation values, we rank the features based on their correlation with the target variable. Features with higher correlation scores are considered more relevant.
- We select a subset of the top-ranked features based on a predetermined threshold or a fixed number of features to be retained.

2. Wrapper Method

- The wrapper method, starting with individual features and gradually expanding to combinations, generates different subsets of features. For example, it might consider subsets like {age}, {income}, {education level}, {gender}, {age, income}, {age, education level}, {age, gender}, and so on.

- Each feature subset is used to train a machine learning model. The model is evaluated using a validation set or cross-validation, and a performance metric like accuracy is computed.
- The subsets are ranked based on their performance metric scores. The wrapper method selects the subset that achieved the highest accuracy.

3. Embedded Method

- One common embedded method is regularization, specifically L1 regularization or Lasso regression, which adds a penalty term to the standard objective function and in essence, performs feature selection by automatically eliminating irrelevant or redundant features.
- For this purpose, fit a Lasso regression model on the training set using the available features (age, income, education level, and gender) to predict the likelihood of purchasing the product.
- Examine the coefficients of the Lasso regression model. Features with non-zero coefficients are considered important and selected for predicting the target variable. Features with zero coefficients are deemed less relevant and can be discarded.
- Evaluate the performance of the Lasso regression model using the testing set. Calculate relevant metrics such as accuracy, precision, recall, or F1-score to assess the predictive power of the model.