Before understanding feature selection techniques it's important to understand why feature selection itself is important. Consider for a prediction task we have two features, it's intuitive that a person with more age is more likely to have higher income, so both of these might convey the same information and using both of them might be redundant. A completely different case could be let's say we want to predict a person's income and among the features provided we are given the information about the average temperature of the area in which the person lives, this information is redundant as its most likely that a person's income will not be affected by the temperature of the area, using such redundant information not only increases computational complexity but can also lead to poor generalisation as the model might overfit on irrelevant patterns or noise. In general we can say that we remove those features that are either highly correlated with each other or those which are irrelevant to the target variable.

Some commonly used feature selection techniques are,

1. Filter Methods: Filter methods refer to those methods that involve feature selection based on statistical means, for example to find out correlation between features we can use correlation matrix and remove features with high correlation. Some other techniques are variance threshold, chi-square test and mutual information.
2. L1 Regularization: L1 regularization involves an additional term of absolute value of weights to the loss function, this means that our model has to decrease the error as well as reduce the value of irrelevant weights essentially making them 0 by the time of convergence.
3. Sequential feature Selection: These models iteratively build models with different feature subsets and add or remove features based on its effect on the performance, if it improves the feature is kept, if it decreases then it is removed. Most commonly used methods are forward elimination and backward elimination.
4. Dimensionality Reduction: Although dimensionality reduction does not necessarily remove the original features, instead it reduces the dimensions of the data preserving the most relevant information, the reduced dimensions might not contain the original features; they are usually a linear combination of the original features. Most commonly used techniques are LDA and PCA.