Beginners guide for feature selection. (By a beginner)

When working with a large dataset, modelling can be time consuming to run because of number of features. It is not uncommon to have hundreds of features for a model. Then it is critical to weed out irrelevant and subpar features. Then the concept of feature selection comes into play. In this article, I shall try to inform about the widely used techniques for feature selection, and demo some of them.

Feature selection is extremely important step for modelling a computationally efficient model. There are a bunch of a technique for this. Let us start by defining the process of feature selection.

Feature selection is the process of selecting a subset of most relevant predicting features for use in machine learning model building.

It helps a model to perform better by weeding out redundant features and features that are not providing much insight. It is economical in computing power because there are fewer features to train on. Results are more interpretable, and it reduces chance of overfitting by detecting collinear features.

Methods:

1. Filter Methods
   1. Not a machine learning approach.
   2. Based on attributes of the feature.
   3. Model agnostic, i.e., performance does not depend on model used.
   4. Should be used for preliminary screening. Detects constant, duplicated, and correlated features.
   5. Usually not the best performance.
   6. Examples:
      1. Univariate selection (ANOVA)
      2. Chi Square
      3. Based on correlation
      4. Based on variance
2. Wrapper Methods
   1. Uses Machine Learning algorithm.
   2. Performance depends on model selected and data underlying.
   3. Usually suggests the optimal feature subset.
   4. Tries different subset of features to figure out optimal features.
   5. Computationally expensive.
   6. Detect interactions between features.
   7. Possibly best performance in terms of feature elimination. Wrappers are terribly slow when it comes to large datasets.
   8. Examples:
      1. Forward Selection
      2. Backward Selection
      3. Exhaustive Search
3. Embedded Methods
   1. Performs feature selection as building the model.
   2. Less computationally expensive than Wrapper methods.
   3. Best of both worlds, more realistic approach.
   4. Examples:
      1. Lasso / Ridge (regularized features in form of ElasticNet)
      2. Tree based selection.
      3. Regression coefficients (features must be standardized).
4. Hybrid Methods
   1. Mixture of all the techniques above.
   2. Less computationally expensive than Wrapper methods.
   3. Examples:
      1. Feature shuffling
      2. Recursive feature elimination
      3. Recursive feature addition

To demo these techniques, I performed feature selection based on correlation, ANNOVA, Forward Selection, RFECV, Lasso techniques on the king county housing dataset. To keep this article short and not diverge from the star here, Feature Selection, I am briefly going through the steps that I took for the data preparation.

1. I dropped duplicates from `id` column.
2. Filled in NaN’s with 0, and other erroneous inputs to 0 in the 'waterfront', 'view', 'yr\_renovated', and 'sqft\_basement' features.
3. Casted appropriate dtypes.
4. Removed outliers from the data.
5. Scaled the data to MinMaxScaler (This scaler scales and translates each feature individually such that it is in the given range on the dataset, e.g., between zero and one) for use in lasso.