DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES

Machine Learning Lab

Master of Science in Signal Theory and Communications

TRACK: Signal Processing and Machine Learning for Big Data

Departamento de Señales, Sistemas y Radiocomunicaciones
E.T.S. Ingenieros de Telecomunicación
Universidad Politécnica de Madrid

Feature selection can be done in multiple ways but there are broadly 3 categories of it:

1. Filter Method
2. Wrapper Method
3. Embedded Method

https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b

1. Filter Method:

As the name suggest, in this method, you filter and take only the subset of the relevant features.

The model is built after selecting the features.

The filtering here is done using correlation matrix and it is most commonly done using Pearson correlation.

Univariate Filters

Another approach to feature selection is to pre-screen the predictors using simple univariate statistical methods then only use those that pass some criterion in the subsequent model steps.

As an example, it has been suggested for classification models, that predictors can be filtered by conducting some sort of *k*-sample test (where *k* is the number of classes) to see if the mean of the predictor is different between the classes. Wilcoxon tests, *t*-tests and ANOVA models are sometimes used.

Predictors that have statistically significant differences between the classes are then used for modeling.

https://topepo.github.io/caret/feature-selection-using-univariate-filters.html

For classification you can try:

AUC Meets the Wilcoxon-Mann-Whitney U-Statistic-

https://blog.revolutionanalytics.com/2017/03/auc-meets-u-stat.html

The area under an ROC curve (AUC) is commonly used in machine learning to summarize the performance of a predictive model with a single value.

But you might be surprised to learn that the AUC is directly connected to the Mann-Whitney U-Statistic, which is commonly used in a robust, non-parametric alternative to Student's t-test.

Here I'll use 'literate analysis' to demonstrate this connection and illustrate how the two measures are related.

Feature: Weight p-value: 8.18200346766231e-07
Feature: Age p-value: 0.003197705106764703
Feature: Cervical p-value: 5.18199378130727e-09
Feature: BMI p-value: 5.029582313553199e-07
Feature: A_Ancho2 p-value: 0.0037616243394976183
Feature: I_Form3 p-value: 0.02979447496083049
Feature: I_Form4 p-value: 0.04598472480960474
Feature: O_Form3 p-value: 0.049349618595081454

Feature: IAH p-value: 2.668948268665901e-30
Feature: Weight p-value: 1.0436965320992271e-06
Feature: Age p-value: 0.0022783299529905107
Feature: Cervical p-value: 1.3995233873782936e-09
Feature: OSA p-value: 8.3741911410896e-40
Feature: BMI p-value: 9.008366766130032e-08
Feature: A_Ancho2 p-value: 0.003501353021712575
Feature: E_Ancho3 p-value: 0.006629051435287094
Feature: I_Form3 p-value: 0.025440484119445113
Feature: O_Form3 p-value: 0.01608439345194229
Feature: U_Form3 p-value: 0.028831724756362354

2. Wrapper Method (I):

- A wrapper method needs one machine learning algorithm and uses its performance as evaluation criteria.

- This means, you feed the features to the selected Machine Learning algorithm and based on the model performance you add/remove the features.

- This is an iterative and computationally expensive process but it is more accurate than the filter method.

2. Wrapper Method (II):

There are different wrapper methods such as Backward Elimination, Forward Selection, Bidirectional Elimination and RFE.

3. Embedded Method (I):

Embedded methods are iterative in a sense that takes care of each iteration of the model training process and carefully extract those features which contribute the most to the training for a particular iteration.

Regularization methods are the most commonly used embedded methods which penalize a feature given a coefficient threshold.

3. Embedded Method (II):

- Feature selection using Lasso regularization. If the feature is irrelevant, lasso penalizes it’s coefficient and make it 0. Hence the features with coefficient = 0 are removed and the rest are taken.

- Random Forest also provides feature importance

- A single tree can also give some information on relevant features