



Cheatsheet:Caret Package

CARET (Classification And Regression Training) is a library in R which provides a set of functions that attempt to streamline the process for creating predictive models.

1.Data Splitting

Function	Description
1 <code>createDataPartition(y,p=0.8)</code>	<code>createDataPartition</code> splits a vector 'y' with 80 percent data in one part and 20 percent in other part <code>createDataPartition(y,p=0.8)</code>
2 <code>maxDissim(a,b,n=2)</code>	It creates subsamples from 'b' which are at a maximum Dissimilarity from 'a' (<code>a,b,n=2</code>)

2.Data Pre-Processing

Function	Description
1 <code>preprocess(x, method=c("center","scale"))</code>	It is used to perform preprocessing tasks like centering, scaling and imputing missing values in a dataset
2 <code>BoxCoxTrans(y,...)</code>	To remove skeweness in a vector by using boxcoxtransformations on it.
3 <code>downSample(x,y,yname="class")</code>	It is used to randomly sample the data so that every class has the same frequency as the minority class.
4 <code>dummyVars(formula,...)</code>	It creates a full set of dummy variables for categorical variables

3.Feature Selection

Function	Description
1 <code>gafs.default(x,y,...)</code>	It is used to perform supervised feature selection using genetic algorithms
2 <code>nearZeroVar(x,...)</code>	It is used to identify predictors that have zero or near zero variance.
3 <code>pickSizeBest(x,metric,maximise)</code>	It is used to perform backward selection
4 <code>rfe(x,...)</code>	It is used to perform a simple backward selection
5 <code>varImp(object,...)</code>	It is used to calculate variable importance for classification and regression models

4.Model Tuning

Function	Description
1 <code>trainControl</code>	It is used for controlling training parameters like resampling, number of folds, iteration etc.
2 <code>oneSE(x,metric,maximise)</code>	This function is used to set tuning parameters of a model.

5.Visualization

Function	Description
1 <code>calibration(x,data)</code>	It is used to draw calibration plot that describe show consistent model probabilities are with the observed event rate.
2 <code>densityplot.rfe(x,data,...)</code>	Lattice functions for plotting resampling results of recursive feature selection
3 <code>featureplot(x,y,plot...)</code>	A shortcut to produce lattice plots
4 <code>plotClassProbs</code>	It is used to plot predicted probabilities in classification model .
5 <code>plotObsVsPred</code>	It is used to plot observed vs predicted results in Classification and Regression Models

For more Infographics log on :
<http://www.analyticsvidhya.com>