



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

2.Data Pre-Processing

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

3.Feature Selection

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

4.Model Tuning

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

5.Visualization

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

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