Human Activity Recognition using Decision Tree

# Data Set:

* The excel file “HAR” is a simplified version of “Human Activity Recognition Using Smartphones Data Set” ([https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#](https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones))
* The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ
* These signals were used to estimate variables of the feature vector for each pattern: ‘-XYZ’ is used to denote 3-axial signals in the X, Y and Z directions
* In the “HAR” dataset, we only use acceleration signal, which was separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityACC-XYZ)
* The set of variables that were estimated from these signals are:
  + Mean(): Mean value
  + Std(): Standard deviation
  + Max(): Largest value in array
  + Min(): Smallest value in array
  + Iqr(): Interquartile range
  + Correlation (): correlation coefficient between two signals
* There are six prediction labels:

1 WALKING

2 WALKING\_UPSTAIRS

3 WALKING\_DOWNSTAIRS

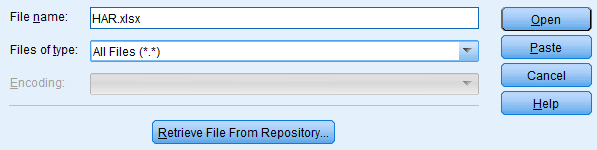
4 SITTING

5 STANDING

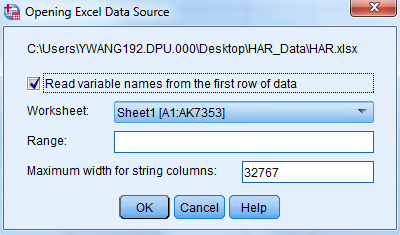
6 LAYING

# SPSS hands-on tutorial on classification

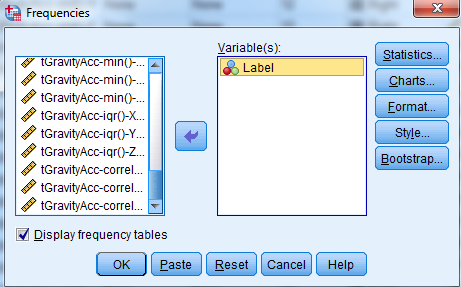
1. Start SPSS Statistics
2. Open the HAR data as an Excel file by clicking “File 🡺 Open 🡺 Data”



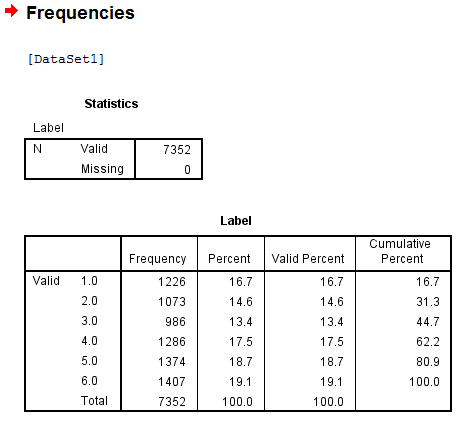
* 1. Make sure that you check the box indicating that the names of the variables show up in the first row of your Excel file



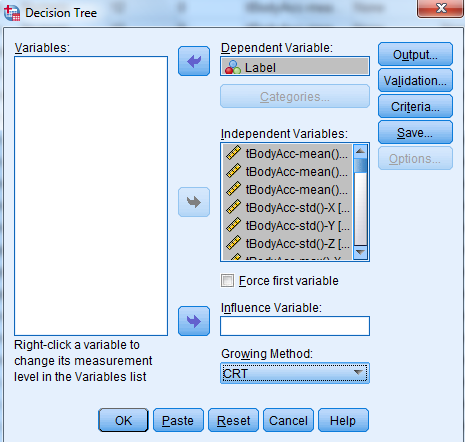
* 1. After opening the data file in SPSS, analyze the data. Run some descriptive statistics on the class variable (the last column in the Excel file) by clicking “Analyze 🡺 Descriptive Statistics 🡺 Frequencies”. Select the “Label” variable and drag it to the right window.



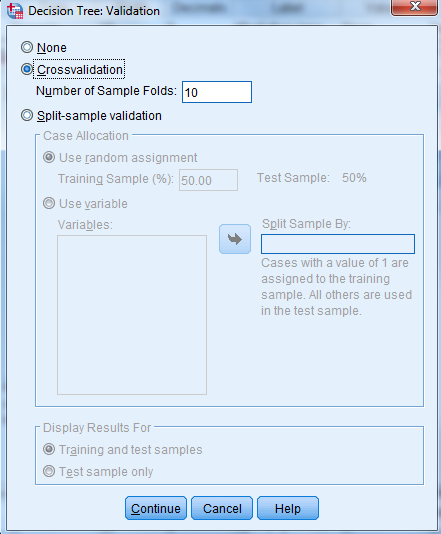
You will see the distribution of each prediction label



1. To model the data using classification, go to ANALYZE, then CLASSIFY, and then choose TREE
2. After that, select C&RT for the growing method for the tree to be built (other options are Quest, CHAID, exhaustiveCHAID)
3. Select all the variables except the label variable and drag them into the Predictors/Independent window (from left window drag them to the right window)
4. Then drag the class variable under the Target/Dependent window



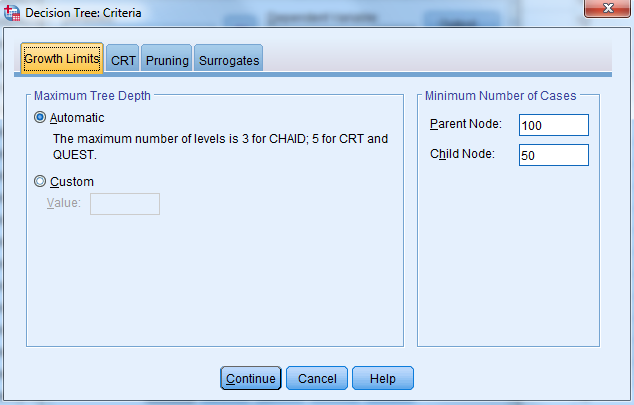
1. Click on Validation button: cross validation or split-sample validation and then click continue the validation window will show up and this will help you with training and testing your decision tree. One of the best way to do it when the dataset is not large enough is to choose Cross Validation



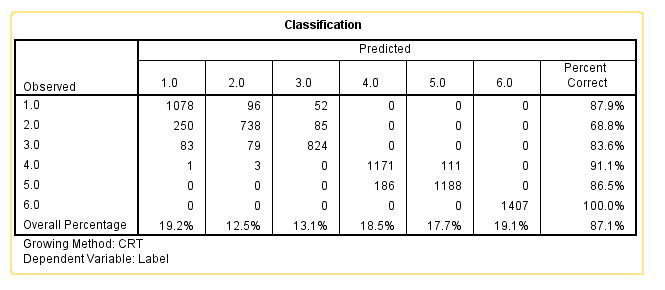
1. Click on Criteria Button:

Three important parameters will show up:

* Depth of the tree (by default =5) which means how many levels you want the tree to have
* Parent node (by default =100) which means how many cases at least you would want to have for a node in order to allow its splitting. For the default value of 100 it means that any node with less than 100 cases will not be split further
* Child node (by default =50) which means how many cases at least you would like to have for the children of a parent node in order to split that parent node further



1. After clicking “continue” and “Ok”, you can see SPSS is running (it may take a while), then you can see the tree (too large to show here) and the misclassification matrix (confusion matrix):



**QUESTIONS:**

1. Your result may not be the exactly same with the result shown above, why? (Assume we used the same data set and the same parameters; Hint: think about how cross- validation works)
2. In your opinion, why some classes are correctly classified more often than others?

**EXERCISE:**

Change the values of the tree parameters and report the misclassification matrix and the accuracy. How did the performance change from the previous tree? In your opinion, why are the results different?