



Demo Decision Tree

Dr. Goutam Chakraborty



Outline

- Start with PVA97NK data and Metadata node using settings from last session
- Add a Replacement node (from Modify tab)
 - Use this node to set all values of DemMedIncome to “missing” if the actual values are less than 1
 - Think **why**?
- Add a data partition node
 - Training 50%, Validation 50% - do we ever use test data?
 - What are some issues in partitioning?
 - Stratification by target variable
 - Easy when the target is ...
 - What about input variables distribution in the training and validation data?



SAS EM Demo Procedure

- Add a decision Tree node (Model tab)
- Run in **Interactive** mode first (via interactive ellipsis button on the properties panel)
 - Right-click the blue box and select **Split Node** from the menu.
 - Select **Edit Rule**. The GiftCnt36 - Interval Splitting Rule dialog box appears.
- Adding more splits
 - Select the lower left partition.
 - Click **Apply**. Then click **OK** twice. The Tree View window is updated to show the additional split
 - Repeat the process for the branch that corresponds to cases with **Gift Count 36 Month** in excess of 2.5.



SAS EM Demo Procedure (continued)

- Changing a Splitting Rule
 - Median Home value best split is at \$67,350
 - Why should we change that?
- Select the node above the label **Median Home Value Region**.
 - Right-click this node and select **Split Node** from the menu.
 - Click **Edit Rule**. The DemMedHomeValue - Interval Splitting Rule dialog box appears
 - Enter **70000** in the **New split point** field.
 - Select **Add Branch**. The Interval Splitting Rule dialog box shows three branches
 - Select **Branch 1** by highlighting the row
 - Click **Remove Branch**
 - Click **OK** twice to close the Interval Splitting Rule dialog box and to update the Tree View window



SAS EM Demo Procedure (continued)

- Creating a Maximal (very large) Tree
 - Select the root node of the tree.
 - Right-click and select **Train Node** from the menu. The tree is grown until *stopping rules* prohibit further growth.
 - Right-click in the gray area behind the tree. Select **View** ⇒ **Fit to page**. The tree view is scaled to provide the general shape of the maximal tree
 - To view the information within each node, right-click in the gray area behind the tree and select **View** ⇒ **Chart tips**. Now the information from each node can be viewed by positioning the mouse pointer over the desired node.
 - Select **View** ⇒ **Subtree Assessment Plot**
 - Save the tree by selecting **File** ⇒ **Save**. Then select **File** ⇒ **Exit** to close the Interactive Tree application