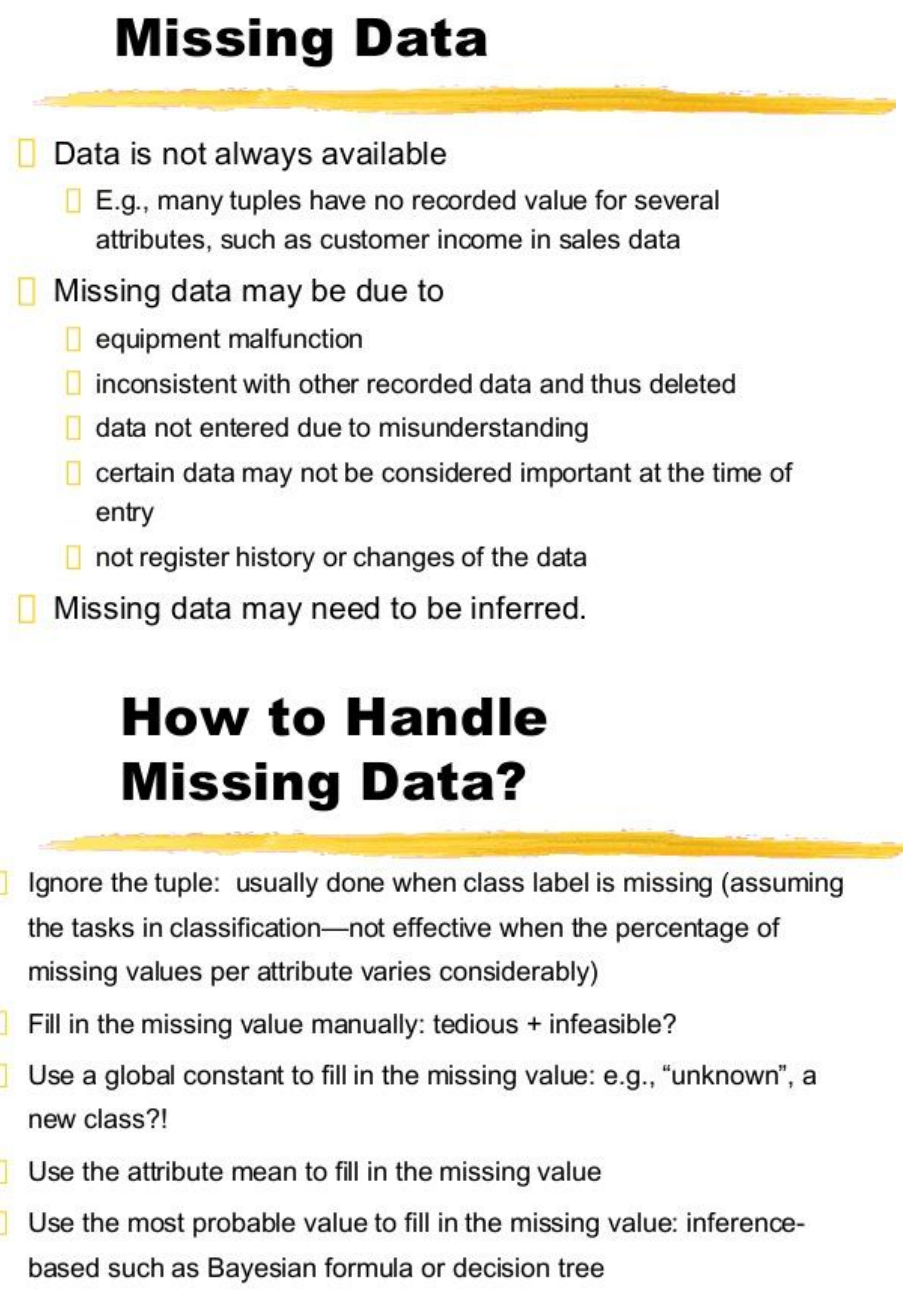
HANDLING MISSING VALUES IN A DECISION TREE



There are several methods used by various decision trees. Simply ignoring the missing values (like ID3 and other old algorithms does) or treating the missing values as another category (in case of a nominal feature) are not real handling missing values. However, those approaches were used in the early stages of decision tree development.

The real handling approaches to missing data does not use data point with missing values in the evaluation of a split. However, when child nodes are created and trained, those instances are distributed somehow.

The following approaches to distribute the missing value instances to child nodes:

1. All goes to the node which already has the biggest number of instances (CART, is not the primary rule)
2. Distribute to all children, but with diminished weights, proportional with the number of instances from each child node (C45 and others)
3. Distribute randomly to only one single child node, eventually according with a categorical distribution (I have seen that in various implementations of C45 and CART for a faster running time)
4. Build, sort and use surrogates to distribute instances to a child node, where surrogates are input features which resembles best how the test feature send data instances to left or right child node (CART, if that fails, the majority rule is used)