* Linear Regression with categorical response variables
  + no longer a straight line
  + Y becomes probability
* Logistic Regression
  + a classification algorithm
  + used to estimate the **probability** of a **qualitative response variable** using a set of quantitative and qualitative variables
* Decision Tree
  + root node
    - the node at the top of the tree
  + leaf nodes (or terminal nodes or leaves)
    - the nodes at the bottom of the tree
* Multi-collinearity
  + situation where two or more predictors are highly correlated
* **Tree Regression**
  + 2 categories = logistic regression
  + e = 2.71828
  + below = 2 types of tree algorithms
    - supervised learning
      * there is a response variable
        + regression problem

response = numeric

linear regression

* + - * + Classification problem

response = categorical

logistic regression

classification tree

* + - unsupervised learning
      * there is no response variable
        + correlation matrix
        + cluster analysis
        + MKT basket analysis
  + knime only implements a classification/decision tree
  + most influencial predictor is the one closest to the top
* **Leaf Nodes**
  + always interpret the leaf nodes
    - to interpret the tree, examine the leaf nodes
  + most important predictor
    - node that initially splits the tree
  + second most important predictor
    - down 1 level
* **Pruning**
  + pruning a tree sacrifices something