

# PROJECT 5 PART 1 (4 POINTS, DUE 3/30)

- Answer the following questions **WITHOUT** the assistance of computer. Show your work.
1. Given the data as shown in the table, find the optimal attribute (A or B) to split, in terms of the GINI index.
  2. Compute the misclassification error for the splitting of A and the splitting of B.
  3. Construct a depth-2 decision tree based on the GINI index.

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

# PROJECT 5 PART 2 (10 POINTS, DUE 4/15)

- Understand the package rpart  
<https://www.rdocumentation.org/packages/rpart/versions/4.1-15>
  - Download the Car Evaluation Data and understand the meaning of each attribute.  
<http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>
1. Describe the trained decision tree model and plot the tree.
  2. Predict the evaluation for the following two types of cars.
    - a) performance car: buying="vhigh", maint="vhigh", doors="2", persons="2", lug\_boot="small", safety="low"
    - b) compact SUV: buying="med", maint="med", doors="4", persons="4", lug\_boot="small", safety="med"
  3. Based on the trained model, what is the criteria for being a very good car?
  4. Split the data into training data (80%) and testing data (20%) . On the same sample data set, compute the training and testing errors for the tree models with depth = 1, 2, 3, ..., 7,8. Construct a table to exhibit your result.
  5. What's your interpretation for the results in Problem 4?
  6. Find the top 3 important factors in the car evaluation. Explain the reasons.

A sample table for Problem 4

depth	train_error	test_error
1	0.05	0.1
2	0.10	0.2
3	0.15	0.3
4	0.20	0.4
5	0.25	0.5
6	0.30	0.6
7	0.35	0.7
8	0.40	0.8