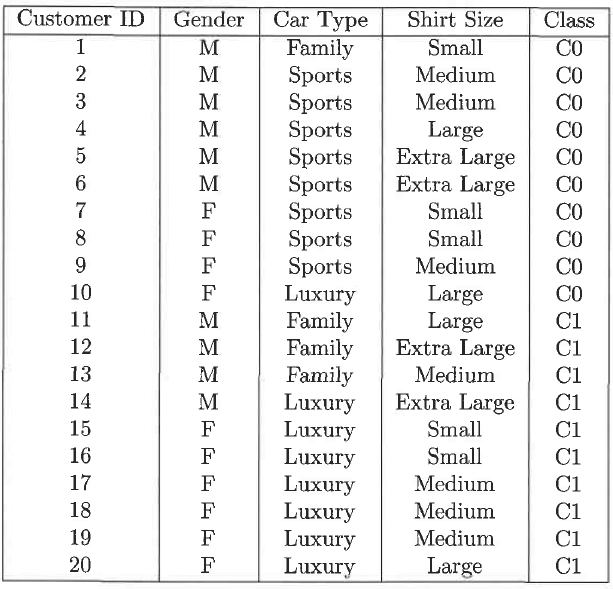
**Lab 9**

**Objectives**

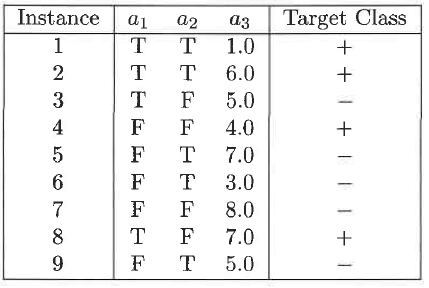
* Be familiar with the classification concepts
* Be familiar with impurity measurement
* Practice building decision tree with best split

**Practice Questions**

1. Consider the training examples shown in the following table for a binary classification problem.



1. Compute the Gini index for the overall collection of training examples.
2. Compute the Gini index for the Customer ID attribute.
3. Compute the Gini index for the Gender attribute.
4. Consider the training example shown in the following table for a binary classification problem.



1. What is the entropy of this collection of training examples with respect to the class attribute?
2. What are the information gains for a1 and a2 relative to these training examples?
3. What is the best split between a1 and a2 according to the information gain?
4. What is the best split between a1 and a2 according to the Gini index?