SC4000/CZ4041/CE4041: Machine Learning
Lecture 5 Tutorial Question Sets

Question 1: Consider the data set shown in Table 1 for a binary classification problem.

Table 1: Data set for Question 1.

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

1. Calculate the information gain when splitting on A and B (using multi-way split on A). Which feature would the decision tree induction algorithm choose?

2. Calculate the gain ratio when splitting on A and B (using multi-way split on A). Which feature would the decision tree induction algorithm choose?