Assignment 2.1 [Hand]

University of San Diego

ADS 502

Dingyi Duan

*Introduction to Data Mining*: Exercises 3.11 – Page 186: Question #3

**3. Consider the training examples shown in Table 3.6 for a binary**

**classification problem.**

**Table 3.6. Data set for Exercise 3.**

**Instance a1 a2 a3 Target Class**

1 T T 1.0 +

2 T T 6.0 +

3 T F 5.0 -

4 F F 4.0 +

5 F T 7.0 -

6 F T 3.0 -

7 F F 8.0 -

8 T F 7.0 +

9 F T 5.0 -

1. **What is the entropy of this collection of training examples with respect to the positive class attribute?**

Graphical user interface, text, application, email

Description automatically generated

1. **What are the information gains of a1 and a2 relative to these training examples?**
2. **For a3, which is a continuous attribute, compute the information gain for every possible split.**
3. **What is the best split (among a1, a2 and a3) according to the information gain?**
4. **What is the best split (between a1 and a2) according to the misclassification error rate?**
5. **What is the best split (between a1 and a2) according to the Gini index?**