3. Consider the training examples shown in below table for a binary classification

problem.

(a) What is the entropy of this collection of training examples with respect

to the positive class?

(b) What are the information gains of *a*1 and *a*2 relative to these training

examples?

(c) For *a*3, which is a continuous attribute, compute the information gain

for every possible split.

(d) What is the best split (among *a*1, *a*2, and *a*3) according to the information

gain?

(e) What is the best split (between *a*1 and *a*2) according to the classification

error rate?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Instance | a1 | a2 | a3 | Target Class |
| 1 | T | T | 1 | + |
| 2 | T | T | 6 | + |
| 3 | T | F | 5 | − |
| 4 | F | F | 4 | + |
| 5 | F | T | 7 | − |
| 6 | F | T | 3 | − |
| 7 | F | F | 8 | − |
| 8 | T | F | 7 | + |
| 9 | F | T | 5 | − |

(f) What is the best split (between *a*1 and *a*2) according to the Gini index?