

2. (4 pt) Consider the task of building a classifier from random data, where the attribute values are generated randomly irrespective of the class labels. Assume the data set contains records from two classes, “+” and “-.” Half of the data set is used for training while the remaining half is used for testing.

(a) Suppose there are an equal number of positive and negative records in the data and the decision tree classifier predicts every test record to be positive. What is the expected error rate of the classifier on the test data?

Suppose we have 100 test instances, 50 positive and 50 negative.

The classifier will classify all 50 positive instances correctly, and will classify all 50 negative instances incorrectly. Hence, the error rate is $50/100 = 50\%$

(b) Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability 0.8 and negative class with probability 0.2.

Answer: 50%. **Why?**

(c) Suppose two-thirds of the data belong to the positive class and the remaining one-third belong to the negative class. What is the expected error of a classifier that predicts every test record to be positive?

Answer: 33%. **Why?**

(d) Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability $2/3$ and negative class with probability $1/3$.

Answer: 44.4%. **Why?**