Mark Gameng

CS 422 – HW 5

Part 1 – Questions, Chapter 4

18. + , - classes with half training and half for testing.
    a. Equal number of + and – instances. Decision Tree predicts every test instance to be positive. Thus, it is expected to misclassify half of the instances so the expected error rate is **0.50**
    b. Predicts positive with 0.8 probability, and negative with 0.2 probability. Equal number of + and -, thus N/2 are true positive and N/2 are true negative. With N being the number of test instances.
       20% of the N/2 true positive can be misclassified as negative. $(1 – 0.8)$
       80% of the N/2 true negative can be misclassified as positive. $(1 – 0.2)$
       So $(0.20*N/2+0.8*N/2)/N = (N/2) / N = 0.5$.
       Thus, the expected error rate is **0.50.**
    c. 2/3 positive and 1/3 negative in the data. Decision tree predicts every test instance to be positive. Thus, it is expected to misclassify 1/3 of the instances so the expected error rate is **1/3.**
    d. Predicts positive with 2/3 probability, and negative with 1/3 probability. 2N/3 are true positive and 1N/3 are true negative.
       1/3 of the 2N/3 true positive can be misclassified as negative. $(1 – 2/3)$
       2/3 of the 1N/3 true negative can be misclassified as positive. $(1 – 1/3)$
       So $(1/3 * 2N/3 + 2/3 * 1N/3) / N = (2N/9 + 2N/9) / N = (4N/9) / N = 4/9$
       Thus, the expected error rate is **4/9.**