Homework 3

Jarod Klion

February 23rd, 2022

**Chapter 5:**

1. **Consider a binary classification problem with the following set of attributes and attribute values:**
   1. **Are the rules mutually exclusive?**
      1. No
   2. **Is the rule set exhaustive?**
      1. No
   3. **Is ordering needed for this set of rules?**
      1. Yes
   4. **Do you need a default class for the rule set?**
      1. Yes
2. Diagram

   Description automatically generated with low confidence
3. **Figure 5.1 illustrates the coverage of the classification rules R1, R2, and R3. Determine which is the best and worst rule according to:**

R1: 12 pos, 3 neg

R2: 7 pos, 3 neg

R3: 8 pos, 4 neg

Total: 29 positive examples, 21 negative examples

* 1. **The likelihood ratio statistic :**
     1. R1:
     2. R2:
     3. R3:
     4. R1 is the best rule, and R3 is the worst rule by the likelihood ratio statistic
  2. **The Laplace measure:**
     1. R1:
     2. R2:
     3. R3:
     4. R1 is the best rule, and R3 is the worst rule by the Laplace measure
  3. **The m-estimate measure (with *k* = 2 and *p*+= 0.58):**
     1. R1:
     2. R2:
     3. R3:
     4. R1 is the best rule, and R3 is the worst rule by the m-estimate measure
  4. **The rule accuracy after R1 has been discovered, where none of the examples covered by R1 are discarded.**
     1. R2:
     2. R3:
     3. R2 is chosen because it has higher accuracy than R3
  5. **The rule accuracy after R1 has been discovered, where only the positive examples covered by R1 are discarded.**
     1. R2: 70%
     2. R3: 60%
     3. R2 is preferred because it has higher accuracy than R3
  6. **The rule accuracy after R1 has been discovered, where both positive and negative examples covered by R1 are discarded.**
     1. R2: 70%
     2. R3: 75%
     3. R3 is preferred because it has higher accuracy than R2

1. **Answer the following probability questions about student smokers.** 
   1. **Suppose the fraction of undergraduate students who smoke is 15% and the fraction of graduate students who smoke is 23%. If one­-fifth of the college students are graduate students and the rest are undergraduates, what is the probability that a student who smokes is a graduate student?**
      1. Given probabilities:

P(S|UG) = 0.15

P(S|G) = 0.23

P(G) = 0.2

P(UG) = 0.8

* 1. **Given the information in part (a), is a randomly chosen college student more likely to be a graduate or undergraduate student?** 
     1. P(UG) > P(G), so more likely to be an undergraduate.
  2. **Repeat part (b) assuming that the student is a smoker.**
     1. , so more likely to be an undergraduate still.
  3. **Suppose 30% of the graduate students live in a dorm but only 10% of the undergraduate students live in a dorm. If a student smokes and lives in the dorm, is he or she more likely to be a graduate or undergraduate student? You can assume independence between students who live in a dorm and those who smoke.**
     1. Given probabilities:

P(D|UG) = 0.1

P(D|G) = 0.3

Needed probabilities:

P(D) =

P(S) =

Conditional independence assumption:

P(DS|UG) =

P(DS|G) =

P(UG|DS) =

P(G|DS) =

* + 1. P(G|DS) > P(UG|DS) so more likely to be a graduate student.

1. **Shape, rectangle

   Description automatically generated with medium confidenceConsider the data set shown in Table 5.1**
   1. **Estimate the conditional probabilities for P (A|+), P (B|+), P (C|+), P (A|−), P (B|−), and P (C|−).**
   2. **Use the estimate of conditional probabilities given in the previous question to predict the class label for a test sample (A = 0, B = 1, C = 0) using the naïve Bayes approach.**
      1. The class label should be ‘+’
2. **Table

   Description automatically generated with low confidence**
3. **Consider the plot shown in Figure 5.2**
   1. **Explain how naïve Bayes performs on the data set shown in Figure 5.2.**
      1. The conditional probabilities for each attribute are the same for both class A and class B, so naïve Bayes will perform poorly on this data set.
   2. **If each class if further divided such that there are four classes (A1, A2, B1, and B2), will naïve Bayes perform better?**
      1. Yes, naïve Bayes will perform better
   3. **How will a decision tree perform on this data set (for the two-class problem)? What if there are four classes?**
      1. A decision tree will perform poorly for the two-class problem as there will be no improvement in entropy after splitting; however, four classes will greatly improve the decision tree’s performance.
4. **Repeat the analysis shown in Example 5.3 for finding the location of a decision boundary using the following information:**
   1. **The prior probabilities are P(Crocodile) = 2 x P(Alligator).**
   2. **The prior probabilities are P(Alligator) = 2 x P(Crocodile).**
   3. **The prior probabilities are the same, but their standard deviations are different; i.e., σ(Crocodile) = 4 and σ(Alligator) = 2.**
      1. For x ≤ 7.625, animals would be classified as crocodiles.

For 7.625 < x < 14.375, animals would be classified as alligators.

For x ≥ 14.375, animals would be classified as crocodiles.

1. **A screenshot of a computer

   Description automatically generated with low confidence**
2. **Consider the one-dimensional data set shown in Table 5.4**
   1. **Classify the data point x = 5.0 according to its 1-, 3-, 5-, and 9-nearest neighbors (using majority vote).**
      1. 1-nearest neighbor: +
      2. 3-nearest neighbor: -
      3. 5-nearest neighbor: +
      4. 9-nearest neighbor: -
   2. **Repeat the previous analysis using the distance-weighted voting approach described in Section 5.2.1.**
      1. 1-nearest neighbor: +
      2. 3-nearest neighbor: +
      3. 5-nearest neighbor: +
      4. 9-nearest neighbor: +
3. **Answer the following questions about neural networks.**
   1. **Demonstrate how the perceptron model can be used to represent the AND and OR functions between a pair of Boolean variables.**
      1. AND:
      2. OR:
   2. **Comment on the disadvantage of using linear functions as activation functions for the multilayer neural networks.**
      1. The disadvantage of using linear functions is that not all functions can be represented as a linear function, so the network will be less expressive.