Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 5

Due: November 5, 2023 (100 points)

Spencer Solomon
YOUR GIT USERNAME
solomon.sp@northeastern.edu

Naive Bayes, Bayes Rule

Question 1

Say the variable X will be 1 if the patient has Parkinson’s and 0 if not. Similarly, the variable Y will be 1 if the test determines the patient has Parkinson’s and 0 if not.

$$P(X|Y) = \frac{P(Y|X)P(Y)}{P(X)}$$

$$P(X = 1|Y = 1) = \frac{P(Y=1|X=1)P(Y=1)}{P(X=1)}$$

From the question, we know that the probability a patient tests positive if they have Parkinson’s, P(Y=1|X=1), is .9. We also know that the probability of any given patient having Parkinson’s, P(X=1), is .1. In order to find P(Y=1) we must calculate the total probability for Y=1. So:

$$P(Y = 1) = P(Y = 1|X = 0) * P(X = 0) +$$

$$P(Y = 1|X = 1) * P(X = 1)$$

$$P(Y = 1) = (.1 * .9) + (.9 * .1)$$

$$P(Y = 1) = .18$$

Thus

$$P(X = 1|Y = 1) = \frac{.9 * .18}{.1}$$

$$P(X = 1|Y = 1) =$$

Question 2

We would like to prove: $\sum_i P(A_i|B) = 1$

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

$$P(A) = \sum_i \frac{P(A|B_i)P(A)}{P(B_i)} P(B_i)$$

$$P(A) = \sum_i P(A|B_i)P(A)$$

$$P(A) = P(A) \sum_i P(A|B_i)$$

$$1 = \sum_i P(A|B_i)$$