Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining — Assignment 5

Due: November 7, 2023 (100 points)

So Man Amanda Au-Yeung
https://github.com/amandaay/CS6220DataMining/tree/main

Naïve Bayes, Bayes Rules

The original performance of acoustic classification for Parkinsons Disease leverages speech recordings from controlled subject responses from variety of questions. The task in the competition was to detect whether or not a person X had Parkinsons disease from a sampling of data. As of 2018, the state of the art classifiers have achieved 90% correct classification on a held out dataset, both for subjects who had Parkinsons and those who did not (at equal rates). So, when classifier Y sees person X, it works correctly 90% of the time.

1. Let's say that we run a clinic. This clinic leverages this classifier, which has 90% accuracy. Also, let us say that we know that our current patient load is that 10% of the population have Parkinsons and 90% of the population do not. Let's also say that we're seeing patient X, and the classification algorithm has detected that they have Parkinson's disease. **What's the probability that indeed X has Parkinson's disease?**

Come up with the numerical solution, and show your written work.

1

$X = \begin{cases} 1 & \text{PK} \\ 0 & \text{No PK} \end{cases}$

correctness

$Y = \begin{cases} 1 \\ 0 \end{cases}$

Given: $P(X=1)=0.1$

$P(X=0)=0.9$

$P(Y=0 \mid X=1)=0.1$

$P(Y=1 \mid X=1)=0.9$

$P(Y=1 \mid X=0)=0.1$

$P(Y=0 \mid X=0)=0.9$

0.1 $X=1$:
- 0.9 $Y=1$ correctly detect PK (0.9)(0.1)
- 0.1 $Y=0$ incorrectly detect pt has PK when it doesn't (0.1)(0.1)

0.9 $X=0$:
- 0.9 $Y=1$ correctly detect pt doesn't have PK (0.9)(0.9)
- 0.1 $Y=0$ incorrectly detects pt doesn't have PK aka pt has PK (0.9)(0.1)

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

$$P(Y=1) = P(Y=1 \mid X=1)\,P(X=1) + P(Y=1 \mid X=0)\,P(X=0)$$
$$= (0.9)(0.1) + (0.1)(0.9)$$
$$= 0.18$$

$$P(X=1 \mid Y=1) = \frac{(0.9)(0.1)}{0.18} = 0.5$$

The Sum of Conditional Probabilities

In class, we reviewed three main rules in Bayesian probability inference:

- Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Bayes Theorem:

$$P(A|B)P(B) = P(B|A)P(A)$$

- Total Probability:

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

A well-known outcome of the three sets of rules is the fact that the sum of all the conditional probabilities equals one.

2. Prove that:

$$\sum_i P(A_i|B) = 1$$

2

$$\begin{aligned} P(B) &= \sum_i P(B|A_i)\, P(A_i) \\ &= \sum_i \frac{P(A_i|B)P(B)}{P(A_i)}\, P(A_i) \\ &= \sum_i P(A_i|B)P(B) \\ &= P(B) \sum_i P(A_i|B) \end{aligned}$$

$$\sum_i P(A_i|B) = 1$$