

DS 501: STATISTICAL AND MATHEMATICAL METHODS FOR DATA SCIENCE

Quiz 1

September 11, 2018

PROBLEM

A biologist is conducting experiments on cells which are either malignant or not. Out of all samples he receives, 10% are malignant. He also found that 5% of all malignant cells contain the protein A. It is also known that 25% of all malignant cells have protein B. For the experiments the biologist carried out, he also discovered that 2% of all malignant cells contain protein C. Today he received a sample which has no protein A, but contains protein B and protein C. Such samples with no protein A but containing protein B and C are rare to find and have a 3% chance of occurring. Protein A is normally contained in 50% cells.

What is the probability that this sample is that of a malignant cell if Naïve Bayes' algorithm is used to compute probabilities?

SOLUTION

(using M for malignancy)

$$P(M) = 0.10$$

$$P(A|M) = 0.05$$

$$P(B|M) = 0.25$$

$$P(C|M) = 0.02$$

$$P(\sim A, B, C) = 0.03$$

Use Bayes' rule:

$$P(M|\sim A, B, C) = P(\sim A, B, C|M)P(M)/P(\sim A, B, C)$$

Use naive Bayes' assumption to compute $P(\sim A, B, C|M)$

$$\begin{aligned} P(M|\sim A, B, C) &= P(\sim A|M)P(B|M)P(C|M)P(M)/P(\sim A, B, C) \\ &= (1-0.05)(.25)(.02)(0.1)/0.03 \\ &= 0.0016 \end{aligned}$$