

# CISC7202 Tools for Machine Learning

**Spring 2020**

Due: Feb 5, 2020

1. Assume the probability of a certain disease is 0.01. The probability of testing positive given that a person is infected with the disease is 0.95 and the probability of testing positive given the person is not infected with the disease is 0.05.

(a) Calculate the probability of testing positive.

(b) Use Bayes’ Rule to calculate the probability of being infected with the disease given that the test is positive.

2. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that the notebook was supplied by A?

1. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?
2. Why is “Naive” Bayes naive?

Programming:

1. Load breast\_cancer datasets for KNN classification
2. Please show the curve for training accuracy and test accuracy when K increases from 1 to 30
3. Please show the curve for training accuracy and test accuracy when K increases from 1 to 50

1a. The probability of testing positive:

p(T+) = 0.01 x 0.95 + (1 – 0.01) x 0.05

= 0.059

The probability of testing positive is 0.059

1b. p(D) = % of a certain disease

p( T+ | D ) = % of testing positive given that a person is infected with the disease

The probability of being infected with the disease given that the test is positive:

p( D | T+ ) =

=

= 16.10%

The probability of being infected with the disease given that the test is positive is 16.10%

2. p(A) = % of notebooks to a school is supply by company A

p( D | A ) = % of notebooks found to be defective of company A

p(D) = % of a notebook was found to be defective

= 0.25 x 0.05 + 0.35 x 0.04 + 0.4 x 0.02

= 0.0345

The probability that the notebook was supplied by A is found to be defective:

p( A | D ) =

=

= 36.23%

The probability that the notebook was supplied by A is found to be defective is 36.23%

3. Defective cartridges 🡪 6

Non-defective cartridges 🡪 24

The probability that all the 3 cartridges are defective:

= 0.4926%

The probability that all the 3 cartridges are defective is 0.4926%

4. Naive Bayes is a classification technique and is based on Bayes theorem of conditional probability, which is considered as “naive” is due to it has strong independence assumptions between the features and means changing the value of one feature has no correlation to the change of value of any other features.