CS444 HW #7

Solutions to this assignment should be submitted through Blackboard as .pdf files.

1. Non-Parametric Learning

   - In order to use a k-nearest-neighbors classifier, it is necessary to pick an appropriate value for k. Describe how you could use leave-one-out cross validation to solve that problem.
   - How will a k-nearest-neighbors classifier behave when k is equal to the size of the training set?

2. Assume that A, B, and C, are three mutually independent random variables, and that $P(A = true) = .4$, $P(B = true) = .3$, $P(C = true) = .9$. Find the probabilities that:

   (a) All three are true.

   (b) Exactly two of the three are true.

   (c) None of the three is true.

3. You work at the airport as a passenger screener. You know the following things:

   (a) One passenger in one hundred tries to sneak a bomb through screening.

   (b) The conditional probability that the alarm will go off, given that the passenger has a bomb is .5.

   (c) The conditional probability that the alarm will go off given that the passenger does not have a bomb is .1.

   The alarm goes off. What is the probability that the passenger has a bomb?

4. In a medical study, 100 patients all fell into one of three classes: Pneumonia, Flu, or Healthy. The following database indicates how many patients in each class had fever and headache. A patient presents with a fever but no headache.

   (a) What probability would a naive Bayes classifier assign to the proposition that a patient has Pneumonia. Show your work.

   (b) What probability would a Bayes' optimal classifier assign to that proposition? (A Bayes' optimal classifier *doesn't* make any independence assumptions about the evidence variables)

Pneumonia		
Fever	Headache	count
T	T	5
T	F	0
F	T	4
F	F	1
total:		10

Flu		
Fever	Headache	count
T	T	9
T	F	6
F	T	3
F	F	2
total:		20

Healthy		
Fever	Headache	count
T	T	2
T	F	3
F	T	7
F	F	58
total:		70