Problem Set 08

1. Confident (3p)

Imagine running a medical test for the fictional disease *Sthgiw* in *Llefretniw* (population = 1,000), where 40% are infected. The test has a false positive rate (FP/C-) of 5% and no false negative rate (FN/C+). Create the confusion matrix. What's the positive predictive value (probability that it correctly indicates an infection if a person receives a positive test)?

Now consider the same test applied in *Nurrevir* (population = 1,000), but where only 2% are infected. Create the confusion matrix and recalculate the probability of actually being infected after one is told that one is infected using the same medical test.

Suddenly we have the new disease *Sreklaw*. Only one in a million people gets this disease. We develop a new test that gives us 99% of the time the correct result (99 percent of the time, it gives true if the subject is infected, and false if the subject is healthy). We give the test to everybody in *Soretsew* (population = 1,000,000). How happy are we with our 99% accurate test?

2. Rule (5p)

We have the following credit card dataset:

Sample	Watches advertisement in/on				Sex
	Magazine	Paper	Internet	TV	
S01	Yes	No	No	Yes	Male
S02	Yes	Yes	Yes	No	Female
S03	Yes	No	No	No	Male
S04	No	Yes	Yes	Yes	Male
S05	Yes	Yes	Yes	No	Female
S06	No	Yes	No	No	Female
S07	No	Yes	Yes	Yes	Male
S08	No	No	No	No	Female
S09	Yes	No	No	No	Male
S10	Yes	Yes	Yes	No	Female

Create the ordered rule list. With that, form a single rule and an unordered rule list.

Classify the following sample according to your rules:

Sample	Watches advertisement in/on				Sex
	Magazine	Paper	Internet	TV	
S11	Yes	No	No	No	?

3. Notation (2p)

Match up the unordered English statements with their associated probability notations and write the probabilities (calculations with Bayes' Theorem and normalization might be needed). If there is no English statement matching a probability, please write one.

English statements
We know that 0.8% of the people have cancer.
If cancer is present, the test returns a correct positive result 98% of the time.
It returns a correct negative result 97% of the time if the cancer is not present.

Probability notation
P(cancer) = ____
P(not cancer) = ____
P(cancer|positive) = ____
P(cancer|negative) = ____
P(not cancer|positive) = ____
P(not cancer|negative) = ____
P(positive|cancer) = ____
P(negative|cancer) = ____
P(positive|not cancer) = ____
P(negative|not cancer) = ____

Deadline:

November 28, 2016 at 8:00 AM