CS 479/679 Pattern Recognition
HW 1 – Dr. Bebis
Solutions

1. After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as it is the probability of testing negative given that you do not have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

T^+: test is positive $\quad T^-$: test is negative

D: has disease $\quad \bar{D}$: does not have disease

Given: $P(T^+/D) = 0.99 \quad P(T^-/\bar{D}) = 0.99$

$P(D) = 0.0001$

We can infer: $P(\bar{D}) = \cancel{0.0001}\ 0.9999$, $P(T^+/\bar{D}) = 0.01$

We need to compute:

$$P(D/T^+) = \frac{P(T^+/D)P(D)}{P(T^+)} \text{ where}$$

$$P(T^+) = P(T^+/D)P(D) + P(T^+/\bar{D})P(\bar{D})$$
$$= 0.99 \times 0.0001 + 0.01 \times 0.9999 =$$
$$= 0.010098$$

So, $$P(D/T^+) = \frac{0.99 \times 0.0001}{0.010098} = 0.0098$$

2. Suppose that *X* and *Y* are discrete random variables. Their joint *pmf* function *P(X,Y)* is given by the table below. Show how to compute P(X), P(Y), P(X/Y) and P(Y/X).

example:

		X		
		x_1	x_2	x_3
Y	y_1	0.2	0.1	0.1
	y_2	0.1	0.2	0.3

$$P(X) = \sum_{i=1}^{n} P(X, y_i) \Rightarrow$$
$$P_X(x_1) = P(x_1, y_1) + P(x_1, y_2) = 0.2 + 0.1 = 0.3$$
$$P_X(x_2) = P(x_2, y_1) + P(x_2, y_2) = 0.1 + 0.2 = 0.3$$
$$P_X(x_3) = P(x_3, y_1) + P(x_3, y_2) = 0.1 + 0.3 = 0.4$$

$$P(Y) = \sum_{i=1}^{n} P(X = x_i, Y) \Rightarrow$$
$$P_Y(y_1) = P(x_1, y_1) + P(x_2, y_1) + P(x_3, y1) = 0.2 + 0.1 + 0.1 = 0.4$$
$$P_Y(y_2) = P(x_1, y_2) + P(x_2, y_2) + P(x_3, y2) = 0.1 + 0.2 + 0.3 = 0.6$$

$$P(X|Y) = \frac{P(x_i, y_j)}{P(Y)} = \begin{array}{|c|c|c|c|}\hline & x_1 & x_2 & x_3 \\ \hline y_1 & 0.2/0.4 & 0.1/0.4 & 0.1/0.4 \\ \hline y_2 & 0.1/0.6 & 0.2/0.6 & 0.3/0.6 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline & x_1 & x_2 & x_3 \\ \hline y_1 & 0.50 & 0.25 & 0.25 \\ \hline y_2 & 0.1\overline{6} & 0.\overline{33} & 0.5 \\ \hline \end{array}$$

$$P(Y|X) = \frac{P(x_i, y_j)}{P(X)} = \begin{array}{|c|c|c|c|}\hline & x_1 & x_2 & x_3 \\ \hline y_1 & 0.2/0.3 & 0.1/0.3 & 0.1/0.4 \\ \hline y_2 & 0.1/0.3 & 0.2/0.3 & 0.3/0.4 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline & x_1 & x_2 & x_3 \\ \hline y_1 & 0.\overline{66} & 0.\overline{33} & 0.25 \\ \hline y_2 & 0.\overline{33} & 0.\overline{66} & 0.75 \\ \hline \end{array}$$

3. This problem investigates the way in which conditional independence relationships affect the amount of information needed in probabilistic calculations.

(a) Suppose we wish to calculate $P(H/E_1,E_2)$, and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculations?

i. $P(E_1, E_2)$, $P(H)$, $P(E_1/H)$, $P(E_2/H)$

ii. $P(E_1,E_2)$, $P(H)$, $P(E_1,E_2/H)$

iii. $P(H)$, $P(E_1/H)$, $P(E_2/H)$

(b) Suppose we know that $P(E_1/H,E_2)=P(E_1/H)$ for all values of H, E_1, E_2. Now which of the above three sets are sufficient?

a) We do not have any independence assumptions:

$$P(H|E_1, E_2) = \left(P(E_1, E_2|H)P(H)\right) \Big/ P(E1, E2)$$

So we need to know these probabilities: $P(H), P(E_1, E_2|H)$ and $P(E_1, E_2)$, which is (set ii).

b) If we know that $\boxed{P(E_1|H, E_2) = P(E_1|H)}$ $*$:

$$\begin{aligned} P(H|E_1, E_2) &= \frac{P(H,E1,E2)}{P(E_1,E_2)} && : P(E_1, H, E_2) = P(E_1|H, E_2)P(H, E_2) \\ &= \frac{P(E_1|H,E_2)P(H,E_2)}{P(E_1,E_2)} && : P(E_1|H, E_2) = P(E_1|H) \; * \\ &= \frac{P(E_1|H)P(E_2,H))}{P(E_1,E_2)} && : P(E_2, H) = P(E_2|H)P(H) \\ &= \frac{P(E_1|H)P(E_2|H)P(H)}{P(E_1,E_2)} \end{aligned}$$

Thus we need to know the following probabilities: $P(E_1|H)$, $P(E_2|H)$, $P(H)$, and $P(E_1, E_2)$; which is (set i).

4. The task of finding and removing apples that house worms before they get to the grocery store is a big problem. Consumers have been shown to react poorly upon finding worms in their apples. To combat this problem. Wormfinder Inc. has developed an amazing new non-intrusive test for worms in apples. This test is, called WormScan has the incredible false negative rate of exactly 0 (i.e., if an apple is declared by WormScan to be free of worms, it is guaranteed to have no worms in it). Unfortunately, such a performance comes with a cost; the false positive rate is 3% (i.e., 3% of all good apples are marked as having a worm inside). Statistically, it has been found that 0.2% (1 in 500 apples) have worms.

- **(a)** What percentage of the apples will test as having worms?
- **(b)** Given that an apple has tested as having worms, what is the probability that there is a worm inside?

T^+: test is positive, T^-: test is negative

W: apple has worm, $\bar{W}$: apple does not have worm

Given: ~~[illegible]~~ $FN = 0$ $\quad P(T^-/w) = 0$ $\quad P(T^+/w) = 1.0$

$FP = 3\%$ $\quad P(T^+/\bar{w}) = 0.03$

$P(w) = 0.002$

(a) $P(T^+) = P(T^+/w)P(w) + P(T^+/\bar{w})P(\bar{w})$

$= 1 \times 0.002 + 0.03 \times 0.998 = 0.03194$

(b) $P(w/T^+) = \dfrac{P(T^+/w)\,P(w)}{P(T^+)} = \dfrac{1 \times 0.002}{0.03194} = 0.0626$