1. (10 points total, 5 pts off for each wrong answer, but not negative)

a. (5 pts) Write down the definition of P(H | D) in terms of P(H), P(D), P(H ∧ D), and P(H ∨ D).

P(H | D) = P(H ∧ D) / P(D)

b. (5 pts) Write down the expression that results from applying Bayes' Rule to P(H | D).

P(H | D) = P(D | H) P(H) / P(D)

c. (5 pts) Write down the expression for P(H ∧ D) in terms of P(H), P(D), and P(H ∨ D).

P(H ∧ D) = P(H) + P(D) − P(H ∨ D)

d. (5 pts) Write down the expression for P(H ∧ D) in terms of P(H), P(D), and P(H | D).

P(H ∧ D) = P(H | D) P(D)

2. (10 pts total, 5 pts each) We have a database describing 100 examples of printer failures. Of these, 75 examples are hardware failures, and 25 examples are driver failures. Of the hardware failures, 15 had Windows operating system. Of the driver failures, 15 had Windows operating system. Show your work.

a. (5 pts) Calculate P(windows | hardware) using the information in the problem.

P(windows | hardware) = P(windows, hardware) / P(hardware) = 0.15 / 0.75 = 0.2

b. (5 pts) Calculate P(driver | windows) using Bayes' rule and the information in the problem.

P(driver | windows) = P(windows | driver) P(driver) / P(windows) = (0.15 / 0.25) * 0.25 / 0.3 = 0.5

3. (5 pts) 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 when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that it is a rare disease, striking only 1 in 10,000 people of your age. What is the probability that you actually have the disease?

P(disease | test) = P(test | disease) P(disease) / P(test)

= P(test | disease) P(disease) / [P(test | disease) P(disease) + P(test | −disease) P(−disease)]

= 0.99 * 0.0001 / [0.99 * 0.0001 + 0.01 * 0.9999]

≈ 0.009804

4. (15 pts total, 5 pts each) Suppose you are given a bag containing n unbiased coins. You are told that $n-1$ *of these coins are normal, with heads on one side and tails on the other, whereas one coin is a fake, with heads on both sides. Show your work for the questions below.*

a. Suppose you reach into the bag, pick out a coin uniformly at random, flip it, and get a head. What is the conditional probability that the coin you chose is the fake coin?

P(fake | heads) = P(heads | fake) P(fake) / P(heads)

= P(heads | fake) P(fake) / [P(heads | fake) P(fake) + P(heads | −fake) P(−fake)]

= 1 * (1/n) / [1 * (1/n) + 0.5 * (n−1) / n]

= 2 / (n+1)

b. Suppose you continue flipping the coin for a total of k times after picking it and see k heads. Now what is the conditional probability that you picked the fake coin?

P(fake | k_heads) = P(k_heads | fake) P(fake) / P(k_heads)

= P(k_heads | fake) P(fake) / [P(k_heads | fake) P(fake) + P(k_heads | −fake) P(−fake)]

= 1 * (1/n) / [1 * (1/n) + 2^{-k} * (n−1) / n]

= 2^k / (2^k+n−1)

c. Suppose you wanted to decide whether a chosen coin was fake by flipping it k times. The decision procedure returns FAKE if all k flips come up heads, otherwise it returns NORMAL. What is the (unconditional) probability that this procedure makes an error on coins drawn from the bag?

The procedure makes an error only if a normal coin is drawn and all *k* flips come up heads.

P(k_heads, −fake) = P(k_heads | −fake) P(−fake) = (n−1) / n2^k

5. (10 pts total, 5 pts each) Consider the learning data shown in Figure 18.3 of your book (both 2nd & 3rd ed.). Your book (Section 18.3, "Choosing Attribute Tests") shows that when considering the root node, Gain(Patrons) ≈ 0.541 while Gain(Type) = 0. Calculate Gain (Alternate) and Gain(Hungry) for the root.

Recall (Section 18.3) that

B(q) = −(q $\log_2$ q + (1−q) $\log_2$ (1−q))

Remainder(A) = $\sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B(\frac{p_k}{p_k+n_k})$

Gain(A) = B(p / (p+n)) – Remainder(A)

a. (5 pts) *Gain (Alternate)* = 1 – [$\frac{6}{12} B\left(\frac{3}{6}\right) + \frac{6}{12} B\left(\frac{3}{6}\right)$] = 0

b. (5 pts) *Gain(Hungry)* = 1 – [$\frac{7}{12} B\left(\frac{5}{7}\right) + \frac{5}{12} B\left(\frac{1}{5}\right)$] ≈ 0.196

6. (15 pts total, 5 pts each) Consider an ensemble learning algorithm that uses simple majority voting among M learned hypotheses (you may assume M is odd). Suppose that each hypothesis has error ε where $0.5 > \varepsilon > 0$ and that the errors made by each hypothesis are independent of the others'. Show your work.

a. (5 pts) Calculate a formula for the error of the ensemble algorithm in terms of M and ε.

The ensemble makes an error just in case (M+1) / 2 or more hypotheses make an error simultaneously. Recall that the probability that exactly k hypotheses make an error is

P(exactly k hypotheses make an error) = $\binom{M}{k} \varepsilon^k (1-\varepsilon)^{(M-k)}$

where $\binom{M}{k}$, read "M choose k," is the number of distinct ways of choosing k distinct objects from a set of M distinct objects, calculated as $\binom{M}{k} = \frac{M!}{k!(M-k)!}$, where x!, read "x factorial," is x! = 1*2*3*…*x. Then,

P(error) = $\sum_{k=(M+1)/2}^{M} P(\text{exactly } k \text{ hypotheses make an error}) = \sum_{k=(M+1)/2}^{M} \binom{M}{k} \varepsilon^k (1-\varepsilon)^{(M-k)}$

b. (5 pts) Evaluate it for the cases where M = 5, 11, and 21 and ε = 0.1, 0.2, and 0.4.

	M=5	M=11	M=21
ε=0.1	0.00856	2.98e-4	1.35e-6
ε=0.2	0.0579	0.0117	9.70e-4
ε=0.4	0.317	0.247	0.174

c. (5 pts) If the independence assumption is removed, is it possible for the ensemble error to be worse than ε? Produce either an example or a proof that it is not possible.

YES. Suppose M=3 and ε = 0.4 = 2/5. Suppose the ensemble predicts five examples e1…e5 as follows.

e1: M1 and M2 are in error, so they out-vote M3 and the prediction of e1 is in error.

e2: M1 and M3 are in error, so they out-vote M2 and the prediction of e2 is in error.

e3: M2 and M3 are in error, so they out-vote M1 and the prediction of e3 is in error.

e4, e5: None of the hypotheses make an error on e4 or e5, so the predictions of e4 and e5 are correct.

The result is that each hypothesis has made 2 errors out of 5 predictions, for an error on each hypothesis of 2/5 = 0.4 = ε, as stated. However, the ensemble has made 3 errors out of 5 predictions, for an error on the ensemble of 3/5 = 0.6 > ε = 0.4.

7. (35 pts total, 5 pts off for each wrong answer, but not negative) Label as TRUE/YES or FALSE/NO.

a. (5 pts) Suppose that you are given two weight vectors for a perceptron. Both vectors, w1 and w2, correctly recognize a particular class of examples. Does the vector w3 = w1 − w2 ALWAYS correctly recognize that same class?

NO. Recall that negating the terms in an inequality requires reversing the inequality, so it is hard to predict what negating only one set of terms will do.

b. (5 pts) Does the vector w4 = w1 + w2 ALWAYS correctly recognize that same class?

YES. w4 is a positive linear combination of w1 and w2, so the outputs and thresholds both sum.

c. (5 pts) Does the vector w5 = cw1 where c = 42 ALWAYS correctly recognize the same class?

YES. w5 is a positive linear transform of w1, so the outputs and thresholds are equivalently transformed.

d. (5 pts) Does the vector w6 = dw2 where d = −117 ALWAYS correctly recognize the same class?

NO. Recall that negating the terms in an inequality requires reversing the inequality. The vector w6 will always be exactly INCORRECT, i.e., its class predictions will be inverted. NOTE: If the answer given recognizes this inversion, and is otherwise correct, then give credit.

e. (5 pts) Now suppose that you are given two examples of the same class A, x1 and x2, where $x1 \neq x2$. Suppose the example x3 = 0.5x1 + 0.5x2 is of a different class B. Is there ANY perceptron that can correctly classify x1 and x2 into class A and x3 into class B?

NO. x3 lies on the line segment connecting x1 and x2, and so it cannot be linearly separated from them. NOTE: You can transform the input space in a non-linear way so that the points are linearly separated in the new space; but there is no perceptron that can correctly classify them as given in the problem.

f. (5 pts) Suppose that you are given a set of examples, some from one class A and some from another class B. You are told that there exists a perceptron that can correctly classify the examples into the correct classes. Is the perceptron learning algorithm ALWAYS guaranteed to find a perceptron that will correctly classify these examples?

YES. The perceptron algorithm will find a linear separator if one exists (see your book Section 18.6.3).

g. (5 pts) An artificial neural network can learn and represent only linearly separable classes.

NO. It has a nonlinear transfer function and can have a nonlinear decision boundary (Section 18.7.3).

h. (5 pts) Learning in an artificial neural network is done by adjusting the weights to minimize the error, and is a form of gradient descent.

YES (see your book Section 18.7.4).

i. (5 pts) An artificial neural network is not suitable for learning continuous functions (function approximation or regression) because its transfer function outputs only 1 or 0 depending on the threshold.

NO. It has a nonlinear transfer function and can learn continuous functions (see your book Section 18.7).