Homework 7 , 8 Solutions:

13.3)

Let V be the statement that the patient has the virus, and A and B the statements that the medical tests A and B returned positive, respectively. The problem statement gives:
P(V) = 0.01
P(A|V) = 0.95
P(A|¬V) = 0.10
P(B|V) = 0.90
P(B|¬V) = 0.05
The test whose positive result is more indicative of the virus being present is the one whose posterior probability, P(V|A) or P(V|B) is largest. One can compute these probabilities directly from the information given, finding that P(V|A) = 0.0876 and P(V|B) = 0.1538, so B is more indicative.
Equivalently, the questions is asking which test has the highest posterior odds ratio P(V|A)/P(¬V|A). From the odd form of Bayes theorem: $\frac{P(V|A)}{P(\neg V|A)} = \frac{P(A|V)}{P(A|\neg V)} \frac{P(V)}{P(\neg V)}$

we see that the ordering is independent of the probability of V, and that we just need to compare the likelihood ratios P(A|V)/P(A|¬V) = 9.5 and P(B|V)/P(V|¬V) = 18 to find the answer.

14.6)

a. (c) matches the equation. The equation describes absolute independence of the three genes, which requires no links among them.

b. (a) and (b). The assertions are the absent links; the extra links in (b) may be unnecessary but they do not assert an actual dependence. (c) asserts independence of genes which contradicts the inheritance scenario.

c. (a) is best. (b) has spurious links among the H variables, which are not directly causally connected in the scenario described. (In reality, handedness may also be passed down by example/training.)

d. Notice that the l → r and r → l mutations cancel when the parents have different genes, so we still get 0.5.

G_{mother}	G_{father}	$P(G_{child} = l \mid \ldots)$	$P(G_{child} = r \mid \ldots)$
l	l	$1 - m$	m
l	r	0.5	0.5
r	l	0.5	0.5
r	r	m	$1 - m$

e. This is a straightforward application of conditioning:
P(Gchild = l) = $\sum_{gm,gf}$ P(Gchild = l|gm , gf)P(gm, gf)

= $\sum_{gm,gf}$ P(Gchild = l|gm , gf)P(gm)P(gf) =

= (1 − m)q2 + 0.5q(1 − q) + 0.5(1 − q)q + m(1 − q)2
= q2 – mq2 + q – q2 + m − 2mq + mq2
= q + m − 2mq

f. Equilibrium means that P(Gchild = l) (the prior, with no parent information) must equal P(Gmother = l) and P(Gfather = l), i.e.,
q + m − 2mq = q, hence q = 0.5.
But few humans are left-handed (x ≈ 0.08 in fact), so something is wrong with the symmetric model of inheritance and/or manifestation. The "high-school" explanation is that the "right-hand gene is dominant," i.e., preferentially inherited, but current studies suggest also that handedness is not the result of a single gene and may also involve cultural factors. An entire journal (Laterality) is devoted to this topic.

18.6)

Note that to compute each split, we need to compute Remainder(Ai) for each attribute Ai, and select the attribute the provides the minimal remaining information, since the existing information prior to the split is the same for all attributes we may choose to split on.
Computations for first split: remainders for A1, A2, and A3 are
(4/5)(−2/4log(2/4) − 2/4log(2/4)) + (1/5)(−0 − 1/1log(1/1)) = 0.800
(3/5)(−2/3log(2/3) − 1/3log(1/3)) + (2/5)(−0 − 2/2log(2/2)) ≈ 0.551
(2/5)(−1/2log(1/2) − 1/2log(1/2)) + (3/5)(−1/3log(1/3) − 2/3log(2/3)) ≈ 0.951
Choose A2 for first split since it minimizes the remaining information needed to classify all examples. Note that all examples with A2 = 0, are correctly classified as B = 0. So we only need to consider the three remaining examples (x3,x4,x5) for which A2 = 1.
After splitting on A2, we compute the remaining information for the other two attributes on the three remaining examples (x3,x4,x5) that have A2 = 1. The remainders for A1 and A3 are
(2/3)(−2/2log(2/2) − 0) + (1/3)(−0 − 1/1log(1/1)) = 0
(1/3)(−1/1log(1/1) − 0) + (2/3)(−1/2log(1/2) − 1/2log(1/2)) ≈ 0.667.
So, we select attribute A1 to split on, which correctly classifies all remaining examples

18.17)

The examples map from [x1,x2] to [x1,x1,x2] coordinates as follows:
[−1, −1] (negative) maps to [−1, +1]
[−1, +1] (positive) maps to [−1, −1]
[+1, −1] (positive) maps to [+1, −1]
[+1, +1] (negative) maps to [+1, +1]
Thus, the positive examples have x1x2 = − 1 and the negative examples have x1x2 = + 1. The maximum margin separator is the line x1x2 = 0, with a margin of 1. The separator corresponds to the x1 = 0 and x2 = 0 axes in the original space—this can be thought of as the limit of a hyperbolic separator with two branches.

19.1)

In CNF, the premises are as follows:
¬Nationality(x,n) ∨ ¬Nationality(y,n) ∨ ¬Language(x,l) ∨ Language(y,l)
Nationality(Fernando,Brazil)
Language(Fernando,Portuguese)
We can prove the desired conclusion directly rather than by refutation. Resolve the first two premises with {x/Fernando} to obtain
¬Nationality(y,Brazil) ∨ ¬Language(Fernando,l) ∨ Language(y,l)
Resolve this with Language(Fernando,Portuguese) to obtain
¬Nationality(y,Brazil) ∨ Language(y,Portuguese)
which is the desired conclusion Nationality(y,Brazil) ⇒ Language(y,Portuguese).