1. [10%] True/False Questions

[1% each no partial credit]

a) F

b) F

c) T (Each node in the decision tree divides the examples into two nonempty sets. Since there are n training examples, the number of nodes from the leaf to the root (i.e., the height) is bounded by n.)

d) T (In the worst case, we can have one Boolean output node with n Boolean parents and an n-dimensional, 2^n-big CPT.)

e) F (The decision rule for Naive Bayes cannot be written as: $\Sigma_i w_i x_i > k$.)

f) F

g) T

h) F

i) F

j) T

2. [20%] Bayes Net

2A [6%]

1. **T** [2%]
2. **T** [2%]
3. **F** [2%]

2B. [7%]

P(B+, H+, L-, Q+, E-)

= P(B+) P(H+) P(L- | H+) P(Q+ | B+, H+, L-) P(E- | Q+) [5%, Deduct 1% for each missing term]

= 0.5 * 0.1 * (1 - 0.3) * 0.4 * (1 - 0.6) [2%, -0.5% for each wrongly replaced value]

= 0.0056.

2C. [7%]

P(B | H+, L-, E+)

= αP(B, H+, L-, E+) [2%]

= α(P(B, H+, L-, Q+, E+) + P(B, H+, L-, Q-, E+)) [1%]

= α(<P(B+) P(H+) P(L- | H+) P(Q+ | B+, H+, L-) P(E+ | Q+),

P(B-) P(H+) P(L- | H+) P(Q+ | B-, H+, L-) P(E+ | Q+)>

+ <P(B+) P(H+) P(L- | H+) P(Q- | B+, H+, L-) P(E+ | Q-),

P(B-) P(H+) P(L- | H+) P(Q- | B-, H+, L-) P(E+ | Q-)>

) [1%]

= α P(H+) P(L- | H+) * (<P(B+) P(Q+ | B+, H+, L-) P(E+ | Q+),

P(B-) P(Q+ | B-, H+, L-) P(E+ | Q+)>

+ <P(B+) P(Q- | B+, H+, L-) P(E+ | Q-),

P(B-) P(Q- | B-, H+, L-) P(E+ | Q-)>)

= β * (<0.5 * 0.4 * 0.6, 0.5 * 0.3 * 0.6> + <0.5 * 0.6 * 0.1, 0.5 * 0.7 * 0.1>) [where β := α P(H+) P(L- | H+)] [1%]

= β * 0.5 * (< 0.4 * 0.6, 0.3 * 0.6> + < 0.6 * 0.1, 0.7 * 0.1>)

= γ * (< 0.24, 0.18> + < 0.06, 0.07>)

[where γ := β * 0.5]

[Note: = α(<0.084, 0.063> + <0.021, 0.0245>)]

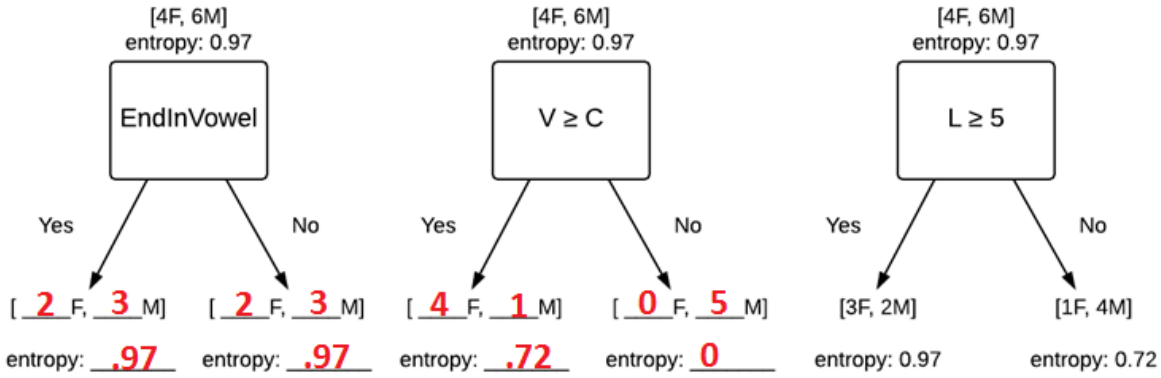
= γ * <0.3, 0.25> [1%, any value pair with the correct ratio (6:5) is ok]

≈ <0.545, 0.455>. [1%, normalization]

[If one of P(B+ | H+, L-, E+) and P(B- | H+, L-, E+) is computed correctly while the other is not, 5% partial credit is given]

3. [23%] Decision Tree Learning

3A. [8%] 2% per correct entropy with number of F/M



3B. [6%] 2% per feature

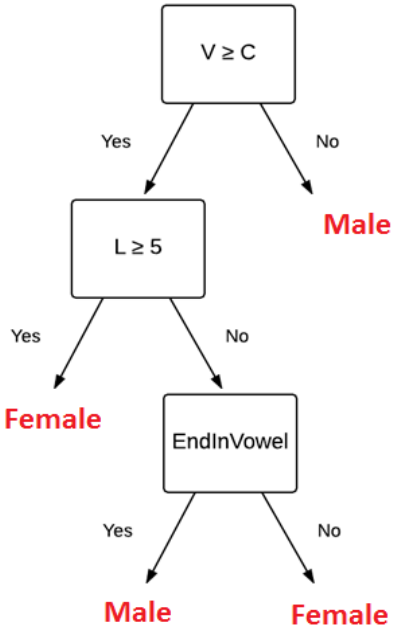
IG (EndInVowel) = 0.97 – (5/10 * .97 + 5/10 * .97)
= 0

IG (V ≥ C) = 0.97 – (5/10 * .72 + 5/10 * 0)
= 0.61

IG (L ≥ 5) = 0.97 – (5/10 * 0.97 + 5/10 * 0.72)
= 0.125

3C. [2%] (V ≥ C) since it has highest information gain.

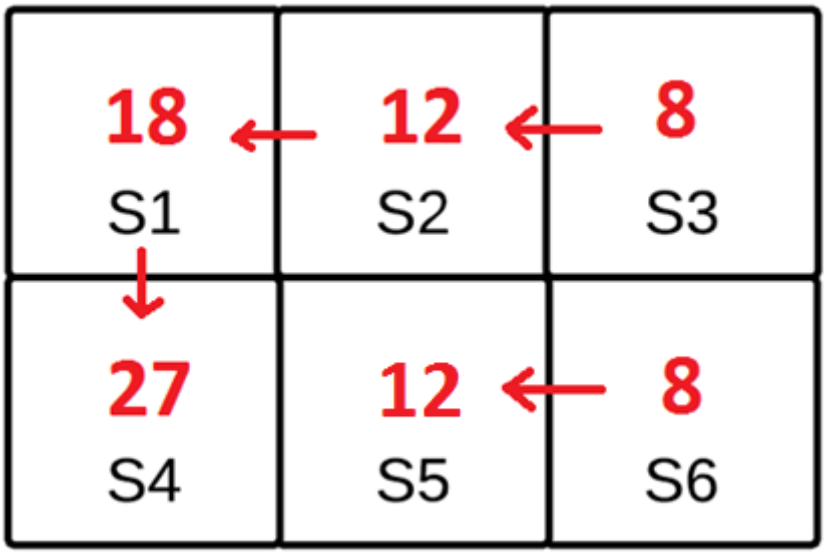
3D. [7%] 1% per correct node



[-0.5%] per leaf that shows [_M, _F]/entropy=0 instead of taking a decision

4. [17%] Markov Decision Process

4A. [7%] 1% per correct V* value (total 6%)
1% for all correct arrows in S1/S2/S3/S6.



4B. [5%] (1% for changed policy – 4% for γ value)
$\gamma^2 R_4 < \gamma R_5$ leads to: S2=>S5 instead of S2=>S1

$\gamma < R_5 / R_4 = 4/9$

4C. [5%]
$V_6^* = p \gamma R_5 + (1-p) \gamma V_6^*$
therefore $V_6^* = p \gamma R / (1- (1-p) \gamma)$ [3% for either form]
$= ¼ * 2/3 * 12 / (1 – ¾ * 2/3)$

$V_6^* = 4$ [2%]

5. [20%] Neural Networks

5A. [4%]

((5 + 1) * 3) + ((3 + 1) * 2) = 18 + 8 = 26
[2% partial credit if the students compute the right number not including the biases: 5 * 3 + 3 * 2 = 21]

5B. [4%]

1. [2%] F.
2. [2%] F. (Overfitting)

5C. [12%]

[6%] Note that $y > 0.5$ if $x > 0$, and $y \leq 0.5$ if $x \leq 0$. Given this, we need to choose w_i so that $w0 + w1 * x1 + w2 * x2$ will be greater than 0 when $x1 \vee x2$ is equal to 1.

X_1	X_2	Y	Constraints on w_i	Credit
0	0	0	$w0 \leq 0$	[1.5%]
0	1	1	$w0 + w2 > 0$	[1.5%]
1	0	1	$w0 + w1 > 0$	[1.5%]
1	1	1	$w0 + w1 + w2 > 0$	[1.5%]

[6%] Similar to previous part, we need to choose w_i so that $w0 + w1 * x1 + w2 * x2$ will be greater than 0 when $x1 \wedge x2$ is equal to 1.

X_1	X_2	Y	Constraints on w_i	Credit
0	0	0	$w0 \leq 0$	[1.5%]
0	1	0	$w0 + w2 \leq 0$	[1.5%]
1	0	0	$w0 + w1 \leq 0$	[1.5%]
1	1	1	$w0 + w1 + w2 > 0$	[1.5%]

6. [10%] Multiple-Choice Questions

[2% per answer]

1.D 2.A 3.A 4.B 5.B