COMP 5970/6970 Assignment 2: 5 questions 5 points 5% Credit
Due before 11:59 PM Friday February 15

Instructions:

1. This is an individual assignment. You should do your own work. Any evidence of copying will result in a zero grade and additional penalties/actions.
2. Enter your answers in this Word file. Submissions must be uploaded **as a single file** (Word or PDF preferred, but other formats acceptable as long as your work is LEGIBLE) to Canvas before the due date and time. Don't turn in photos of illegible sheets. **If an answer is unreadable, it will earn zero points.** Cleanly handwritten submissions (print out this assignment and write answers in the space provided, with additional sheets used if needed) scanned in as PDF and uploaded to Canvas are acceptable.
3. **Submissions by email or late submissions (even by minutes) will receive a zero grade.** No makeup will be offered unless prior permission to skip the assignment has been granted, or there is a valid and verifiable excuse.

Multiple Choice Questions (5 points)

In the following questions, circle the correct choice. If more than one answer is correct, circle all that apply. In those cases, partial credit will be given to partially correct answers. No explanation needed. Incorrect answers or unanswered questions are worth zero points.

1. "Decision tree can represent any function *f*: X → Y." The statement is:

[a] True
[b] False

Answer: [b]

2. "We never test the same attribute twice along one path in a decision tree, but can test it in a different path." The statement is:

[a] True
[b] False

Answer: [a]

Answer the following questions using the following sample training dataset:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

3. What is the sample entropy of the above training dataset?

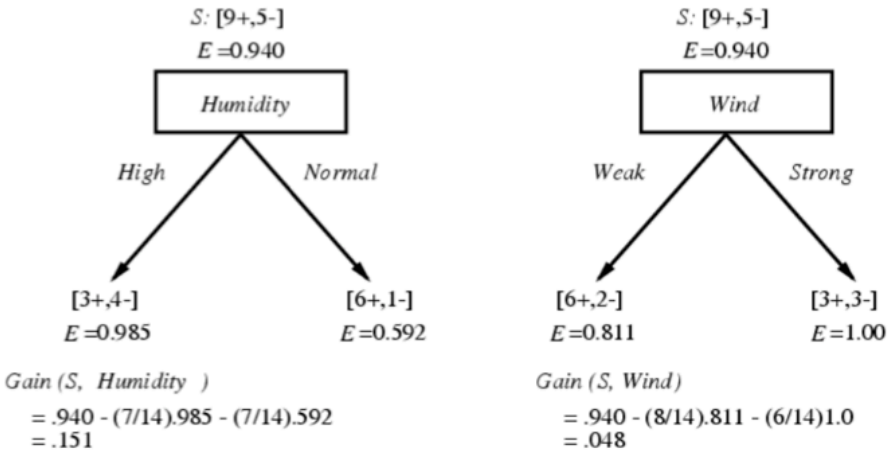
[a] 0.981
[b] 0.893
[c] 0.940
[d] none of the above

Answer: [c]

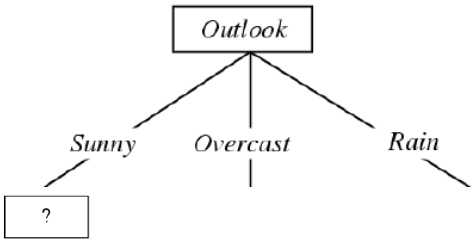
4. In ID3 decision tree learning algorithm, we choose the attribute with the highest information gain. Say you are trying to choose between *Humidity* and *Wind* as the first decision node in the tree. Which one will you choose?

[a] *Humidity*
[b] *Wind*

Answer: [a]

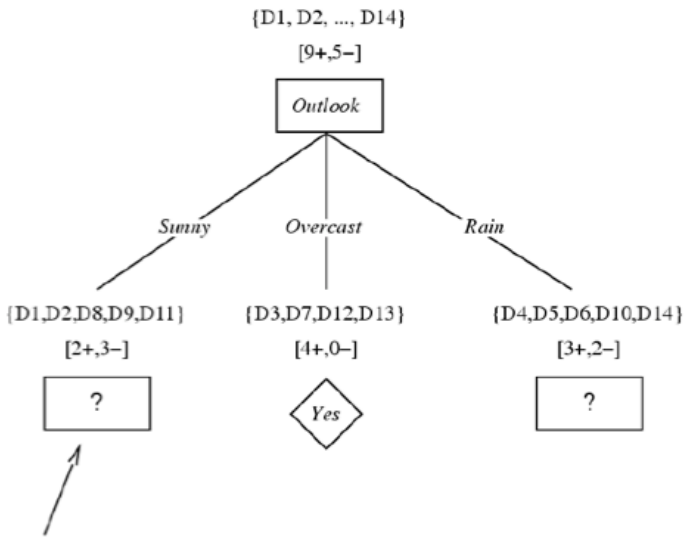


5. Suppose during the greedy top-down growth of ID3 decision tree learning algorithm, we have chosen *Outlook* to be the first decision node (see below tree). Which attribute will you choose at the '?' node?



[a] *Temperature*
[b] *Humidity*
[c] *Wind*
[d] *Outlook*

Answer: [b]



S_{sunny} = {D1,D2,D8,D9,D11}

Gain (S_{sunny}, *Humidity*) = .970 − (3/5) 0.0 − (2/5) 0.0 = .970

Gain (S_{sunny}, *Temperature*) = .970 − (2/5) 0.0 − (2/5) 1.0 − (1/5) 0.0 = .570

Gain (S_{sunny}, *Wind*) = .970 − (2/5) 1.0 − (3/5) .918 = .019