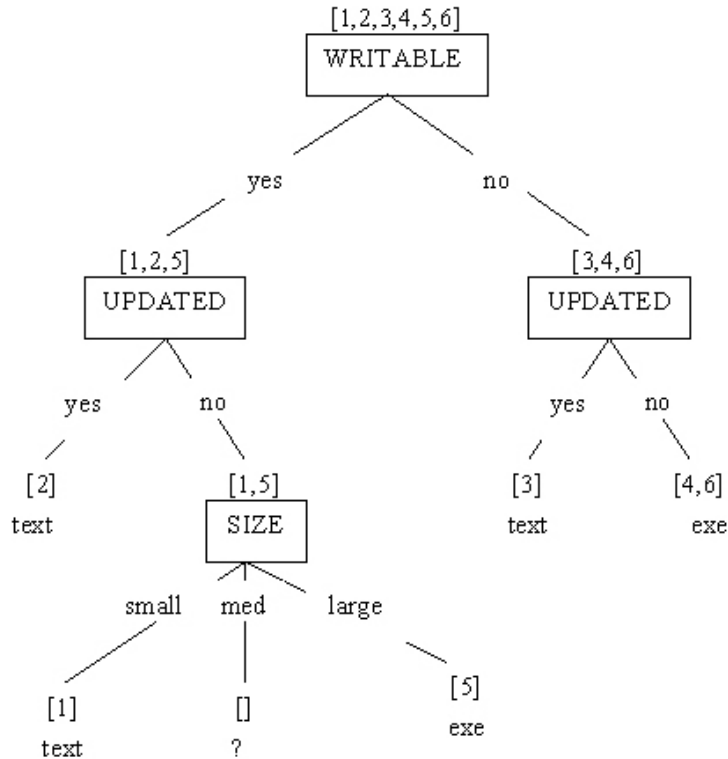# Question 1

1. Compute the decision tree generated by first considering WRITABLE, then UPDATED, then SIZE.

2. What tree is generated by using ID3?
Starting with $s = \{1, 2, 3, 4, 5, 6\}$, we have:

$$
\begin{aligned}
Entropy(s) &= 1.0 \\
Entropy(s|WRITABLE) &= 0.9183 & Gain(s|WRITABLE) &= 0.0817 \\
Entropy(s|UPDATED) &= 0.54 & Gain(s|UPDATED) &= 0.46 \\
Entropy(s|SIZE) &= 0.7925 & Gain(s|SIZE) &= 0.2075
\end{aligned}
$$

Therefore, the best choice is UPDATED. This produces the following partition:

$$
s_1 = \{2, 3\}, \quad s_2 = \{1, 4, 5, 6\}
$$

we have

$$
\begin{aligned}
Entropy(s_1) &= 0 \\
Entropy(s_2) &= 0.811 \\
Entropy(s_2|SIZE) &= 0 & Gain(s_2|SIZE) &= 0.811 \\
Entropy(s_2|WRITABLE) &= 0.5 & Gain(s_2|WRITABLE) &= 0.311
\end{aligned}
$$

Therefore, the best choice is SIZE.
3. How would these trees classify the following two instances:
Tree 1: 1 — text 2 —exe
Tree 2: 1 — text 2 — text

[1,2,3,4,5,6]

UPDATED

yes   no

[2,3]   [1,4,5,6]

text   SIZE

                   small   med   large

                   [1]   [4]   [5,6]

                   text   exe   exe