

1. Giả sử cần dự đoán giá cổ phiếu của 1 công ty. Cây quyết định cần dùng là:

a. Cây phân loại

b. Cây hồi quy

c. Cây gì cũng được

2. Trong quá trình xây cây, khi một node có tất cả mẫu thuộc cùng 1 nhãn thì sẽ:

a. Tìm thuộc tính tốt nhất để tách

b. Chọn node đó làm node lá

c. Quay lại node cha để tìm thuộc tính khác

d. Gộp node đó cùng với các node khác thành 1 node lá

3. ID3 và CART là thuật toán có dạng:

a. Tham lam

b. Quay lui

c. Quy hoạch động

d. Chia để trị

4. Trong cây quyết định, node lá dùng để:

a. Thể hiện nhãn hoặc giá trị dùng để dự đoán

b. Chứa điều kiện để tách

c. Thể hiện độ sâu của cây

d. Thể hiện mức độ quan trọng của 1 thuộc tính

5. Câu nào sau đây đúng:

a. Cây sâu hơn luôn làm tăng độ chính xác trên tập train

b. Cây sâu hơn luôn làm tăng độ chính xác trên tập dữ liệu mới chưa được nhìn thấy

c. Nếu cây sâu hơn cải thiện độ chính xác trên tập train thì nó cũng sẽ cải thiện hiệu suất trên tập dữ liệu mới chưa được nhìn thấy

6. Một thuộc tính rời rạc có thể được chọn nhiều lần trong 1 đường đi từ node gốc đến node lá.

a. Đúng

b. Sai

7. Cây quyết định thường bị underfitting?

a. Đúng

b. Sai

c. 50/50

8. Tiêu chí để lựa chọn thuộc tính trong thuật toán CART?

a. Mean Squared Error (MSE)

b. Gini

c. Entropy

d. Information gain

9. Thuộc tính "gender" có 100% "female" có nhãn "yes". Giá trị gini của "gender" sẽ là:

a. 0

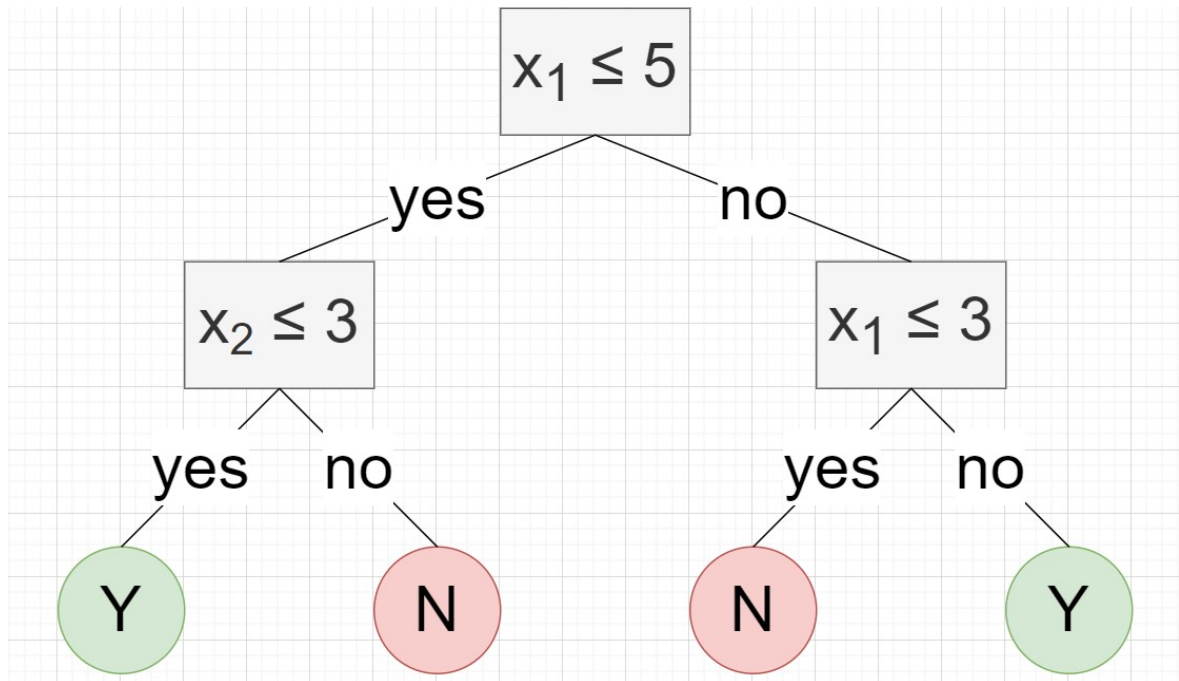
b. 0.25

c. 0.5

d. 0.75

e. 1

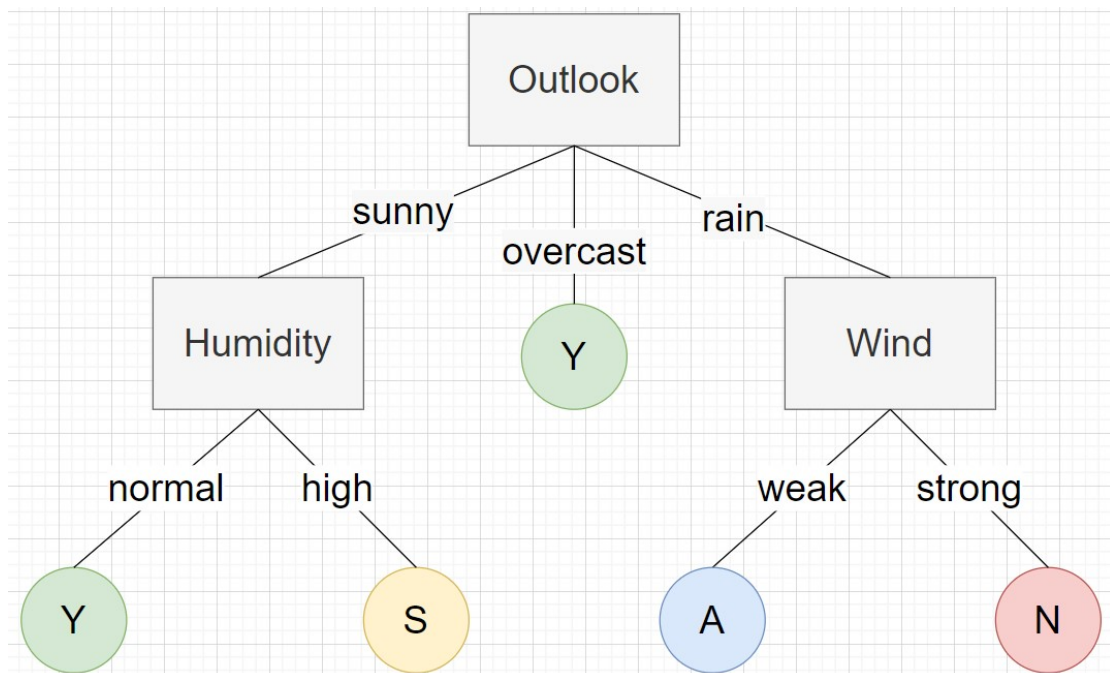
10. Cây sai có hợp lệ không?



a. Có

b. Không

11. Nếu input là {"wind": "strong", "outlook": "sunny", "humidity": "high", "temperature": "cool"} thì output là:



a. A

b. Y

c. S

d. N