1. Mô tả cách thực hiện classification sử dụng decision trees.

Với decision tree ta sẻ học ra một cây nhị phân. Khi sample đi qua các nút ta sẽ so sánh một jhay nhiều feature của sample đó với một threshold (có được khi tranning), tùy vào kết quả mà ta chọn đi tới nút nào tiếp theo. Khi sample đi tới các nút lá thì class của nút đó chính là class của sample.

1. CART algorithm được dùng để làm gì? Mô tả các bước chính của nó.

Thuật toán CART được dùng để xây dựng decision tree dưới dạng binary tree.

Thuật toán sẻ chọn một feature của sample và tính threshold sau đó sẻ tạo ra hai nút và tiếp tục cho tới khi gặp điều kiện dừng. Feature và threshold sẻ được chọn dựa theo cost function.

1. Gini impurity và entropy là gì?

Gini impurity là độ pha tạp của một nút trong decision trees, được tính theo tỷ lệ giữa các sample có class khác class của nút với tổng số sample đi qua nút.

Entropy cũng là công thức tính độ pha tạp nhưng có thể giúp decision tree cân bằng hơn trong một số trường hợp tuy nhiên tốc độ không nhanh bằng Gini.

1. Mô tả cách thực hiện regression sử dụng decision trees.

Trong regression, tại các nút có thuộc tính value thay cho class ở classification và sử dụng Mean Square Error thay vì impurity. Để thực hiện, ta cho sample đi qua các nút khi sample dừng lại ở một nút là thì value của nút đó sẻ là kết quả.

1. Nêu cách thực hiện regularization với decision tree.

Để thực hiện regularization ta chỉ cần giới hạn độ lớn của cây thông qua các tham số sau:

* Max\_depth: giới hạn độ sâu của cây.
* Min\_samples\_split: giới hạn số sample tối thiểu của một nút
* Min\_samples\_leafs: số nút tối thiểu của cây

1. Nếu sử dụng decision tree (không dùng regularization) thì model được tạo có độ sâu khoảng bao nhiêu khi trained với 1000 samples (sử dụng CART algorithm)?   
   Gợi ý: cây nhị phân với k leaf nodes sẽ có độ sâu bao nhiêu?

Khi đó số nút lá trong cây sẻ đạt tối đa bằng số sample vì model bị overfitting. Khi đó độ sâu của cây là log(1000) = 10

1. Nếu một decision tree bị overfitting thì ta nên tăng hay giảm max\_depth của nó? Tại sao?

Ta sẻ giảm max\_depth của nó vì khi bị overfitting model sẻ có rất nhiều nút là và mổi nút gần như chỉ fit với một sample.

1. Nếu một decision tree bị underfitting, thì ta có nên thực hiện feature scaling? Tại sao?

Không, vì feature scaling không ảnh hưởng tới decision tree, do tại mổi nút ta chỉ có 1 threshold cho 1 feature cụ thể, các feeature không liện quan tới nhau.