

Random Forest

Rừng ngẫu nhiên là một thuật toán học máy mà nó kết hợp nhiều cây quyết định (decision tree) để tạo ra kết quả dự đoán chính xác và ổn định hơn.

Cách thuật toán này hoạt động là mình sẽ:

- Xáo trộn các mẫu dữ liệu trong tập huấn luyện để tạo thành các tập con.
- Các tập con ấy được đưa vào thuật toán cây quyết định để tạo ra các cây. Mỗi cây quyết định được huấn luyện một cách độc lập với mỗi tập con.
- Cuối cùng, để đưa ra kết quả dự đoán của thuật toán, các dữ liệu mới ấy sẽ đi qua từng cây quyết định và mỗi cây đưa ra các dự đoán khác nhau:
 - o Đối với bài toán phân loại (classification), bài toán sẽ chọn nhãn xuất hiện nhiều nhất trong số các cây tồn tại làm kết quả dự đoán.
 - o Đối với bài toán hồi quy (regression), bài toán sẽ lấy giá trị trung bình của các dự đoán từ các cây làm kết quả dự đoán.