**Phương pháp tập hợp ensemble**

1. Nội dung:

Là kỹ thuật kết hợp nhiều cây quyết định để tăng hiệu quả dựa đoán so với một cây quyết định duy nhất.

Xây dựng tập hợp các mô hình cơ sở dựa trên tập train.

Kết hợp các mô hình khi phân loại cho độ chính xác cao.

Dựa trên cơ sở bias/ variance

Áp dụng cho nhiều giải thuật cơ sở như cây quyết định, SVM, Naive Bayes

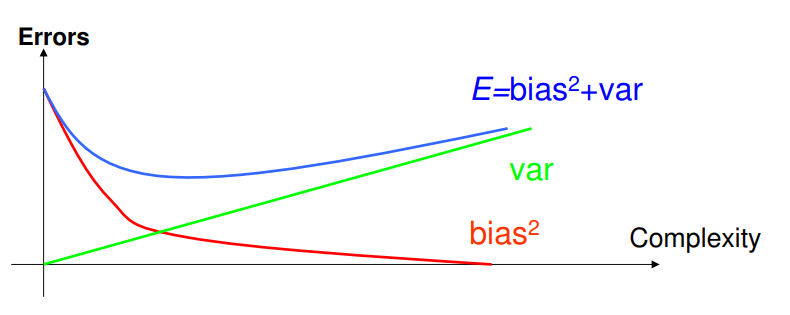
Giải quyết các vấn đề về phân loại, hồi quy, phân cụm,..

Cho kết quả tốt, tuy nhiên không thể dịch được kết quả sinh ra

Được ứng dụng thành công trong các lĩnh vực tìm kiếm thông tin, nhận dạng, phân tích dữ liệu

\*Hiệu quả giải thuật học:

- Bias: thành phần lỗi độc lập với train data

- Variance: thành phần lỗi do biến động liên quan đến sự ngẫu nhiên của train data

\* Phân loại:

+ Averaging technique:

- xây dựng tập hợp các mô hình độc lập nhau

- kết hợp sự phân loại của các mô hình

- Giảm Variance

- Bao gồm: Bagging và Random forest

+ Boosting technique:

- xây dựng tập hợp các mô hình cở sở tuần tự (tập trung vào các lỗi sinh ra từ các mô hình trước)

- Giảm bias