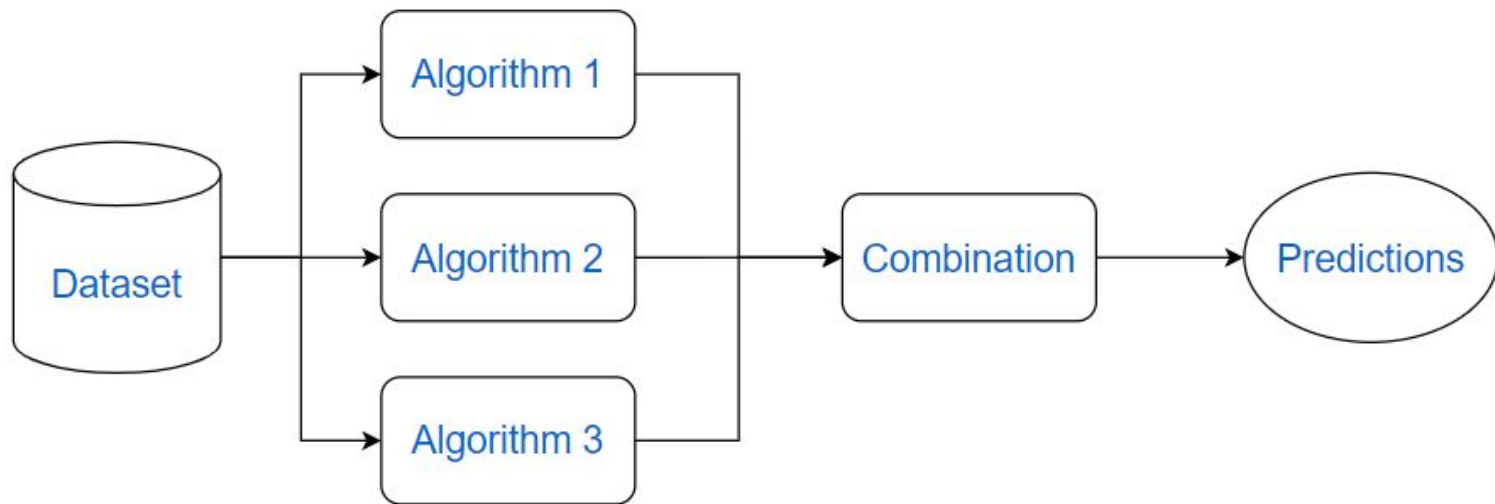


Ensembling



Nội dung

1. Voting ensembles
2. Correlation
3. Weighing
4. Averaging
5. Rank Averaging

1 - Voting ensembles

Tại sao ensemble giúp giảm thiểu error rate?

Model A

1110110011

Model B

1010110011

Với 2 kết quả có được từ model A và B đã được training trên cùng dataset - đâu sẽ là cách tốt nhất để có được điểm số cao nhất trên scoreboard?

1 - Voting ensembles

Tại sao ensemble giúp giảm thiểu error rate?

Model A

1**1**10110011

Model B

1**0**10110011

Model C

1**1**10110011

Label: 1**1**10110011

Predict:

- Model A: 1**1**10110011
- Model B: 1**0**10110011
- Model C: 1**1**10110011

Majority vote:

=> 1**1**10110011

1 - Voting ensembles

Tại sao ensemble giúp giảm thiểu error rate?

Giả sử chúng ta có 3 model A, B, C xử lý tác vụ binary classification với độ chính xác đều là 70%.

Sau khi sử dụng majority vote (voting) 3 model A, B, C thì độ chính xác tăng từ **70%** lên **78,4%** ($0.343 + 0.441 = 0.784$)



Dự đoán cả 3 đều đúng:

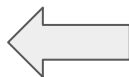
$$0.7 * 0.7 * 0.7 = 0.343$$

1 Đúng - 2 Sai

$$0.3 * 0.3 * 0.7 + 0.3 * 0.7 * 0.3 + 0.7 * 0.3 * 0.3 = 0.189$$

2 Đúng - 1 Sai

$$0.7 * 0.7 * 0.3 + 0.7 * 0.3 * 0.7 + 0.3 * 0.7 * 0.7 = 0.441$$



Dự đoán cả 3 đều sai

$$0.3 * 0.3 * 0.3 = 0.027$$

2 - Correlation

Model A

1111111100
80%

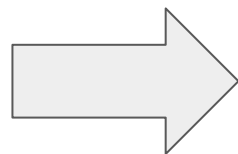
Model B

1111111100
80%

Model C

1011111100
70%

Bài toán dự đoán chuỗi string full “1”



1111111100
80%

Khi 3 model với highly correlated thì
kết quả sau khi sử dụng majority vote
không thấy rõ sự cải thiện.

2 - Correlation

Model A

1111111100
80%

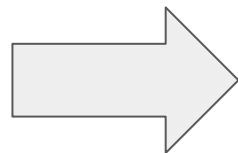
Model B

0111011101
70%

Model C

1000101111
60%

Bài toán dự đoán chuỗi string full “1”



1111111101
90%

Khi 3 model với lower-correlated thì
sau khi sử dụng majority vote cho
thấy kết quả cải thiện rõ rệt.

3 - Weighing

Dựa trên majority vote nhưng sẽ đánh trọng số lên model tốt nhất
=> Weighing majority vote

Model A

80%

Model B

70%

Model C

68%

Weighing majority vote:
 $[n * \text{pred}(A), \text{pred}(B), \text{pred}(C)]$

Với $n > 2$

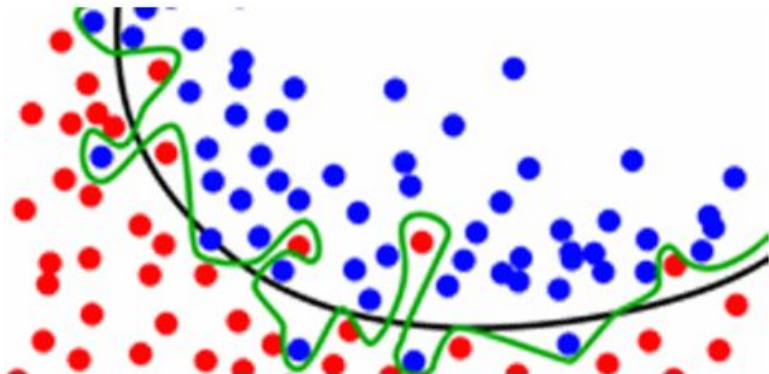
=> Nâng cao sự “uy tín” của best model trong quá trình vote.

4 - Averaging

Averaging hiệu quả cho

- Bài toán classification và regression
- Metrics (AUC, squared error, log-loss score)

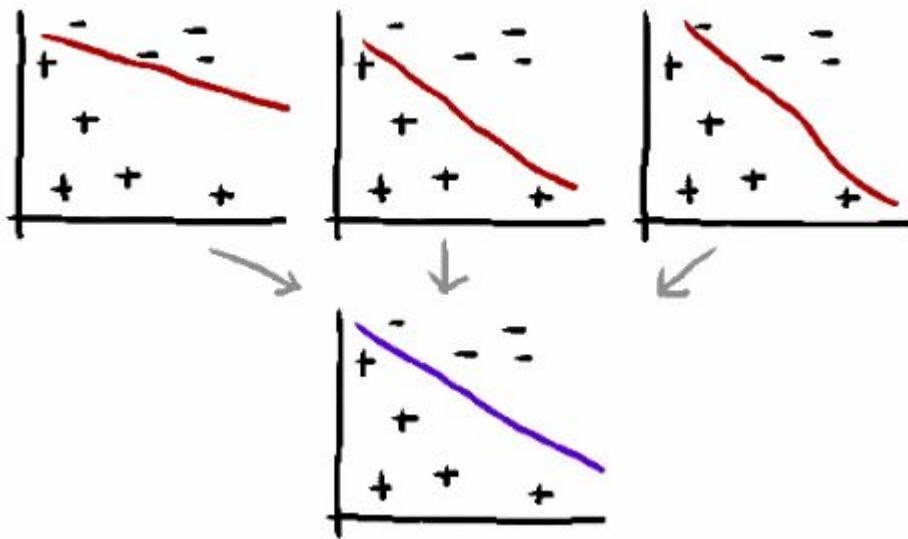
Averaging được tính bằng cách: tính mean cho mỗi model predictions - giúp giảm thiểu overfit.



Sau khi training, model cho ra kết quả mang đến khả năng cao overfit (đường màu xanh) khi model quá cố gắng học các điểm dữ liệu nhiều. Đường màu đen giúp giảm sự ảnh hưởng của các điểm nhiễu này

4 - Averaging

Bag of Words Meets Bags of Popcorn: dựa vào base model để tạo thêm 4 model với `random_weight` sau đó thực hiện tính trung bình trên 5 model đó.



5 - Rank averaging

Một số vấn đề với averaging:

- Không phải các prediction đều được hiệu chỉnh hoàn hảo (Model có thể over/under confident).
- Các dự đoán nhiễu loạn xung quanh 1 data point.

Model A

Id	Prediction
1	0.35000056
2	0.35000002
3	0.35000098
4	0.35000111

Đây có thể là kết quả tốt khi evaluation metric sử dụng ranking hoặc threshold như AUC.

Model B

Id	Prediction
1	0.57
2	0.04
3	0.96
4	0.99

Tuy nhiên khi average với model B thì kết quả ensemble không giúp cải thiện kết quả

5 - Rank averaging

Model A

Id	Rank	Prediction
1	1	0.35000056
2	0	0.35000002
3	2	0.35000098
4	3	0.35000111

Model B

Id	Rank	Prediction
1	1	0.57
2	0	0.04
3	2	0.96
4	3	0.99

Id	Prediction
1	0.33
2	0.0
3	0.66
4	1.0

Kết quả được chuẩn hóa từ 0 đến 1, và kết quả nhận được là một phân phối đều