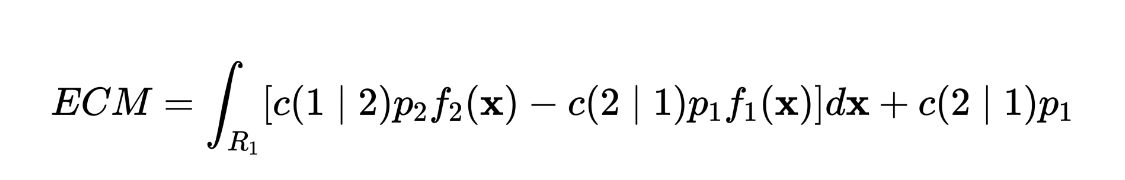
08/03/2022 SEMINAR

1. Trong bài toán phân lớp, quá trình cực tiểu sai số ECM và chuyển ECM từ 2 miền thành 1 miền, Tại sao cần tích phân trong ECM <= 0?



Vì ’số cá thể lớp 1 bị nhầm vào lớp 2’ >= 0

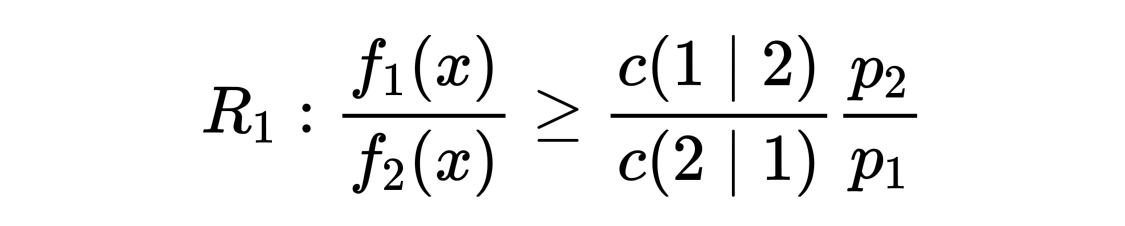
nhân với ‘số cá thể lớp 1’>0

=> c(2|1) p\_1>=0 : không âm

Để cực tiểu sai số ECM thì ECM = 0, suy ra ’tích phân miền R\_1’ = 0 - không âm = không dương

suy ra ’tích phân miền R\_1’ <=0

1. Làm thế nào đề ra được tiêu chí phân lớp như hình dưới? Tiêu chí này gán nhãn là lớp R\_2 được hay không?



Lớp gồm ’gán nhãn đúng’ và ‘bị nhầm thành’.

pt ⇔ f1 \* c(2|1) \* p1 >= f2 \* c(1|2) \* p2

Vế trái: ‘gán nhãn đúng 1’ và ‘lớp 1’ và ‘1 bị nhầm thành 2’.

Vế phải: ‘lớp 2‘ và ’gán nhãn đúng 2’ và ‘‘2 bị nhầm thành 1’.

VT >= VP đồng nghĩa tỉ lệ nghiêng về lớp 1 nhiều hơn và ngược lại.

Cho nên tiêu chí trên KHÔNG ĐƯỢC gán nhãn là lớp R\_2.

1. Vì sao ở bài toán phân lớp 2 quần thể đa biến (con cá có nhiều thuộc tính), thì tiêu chí phân lớp chuyển từ thành ? Nó có lợi gì cho mình sau này?

Bởi vì mục tiêu đạt được hiệu quả trong việc phân lớp, người ta mong muốn 2 điều:

Một, mức độ tập trung của các cá thể trong 1 lớp là lớn (tức phương sai đủ nhỏ).

Hai, 2 lớp phải khác biệt rõ ràng để tránh nhầm lớp (khoảng cách 2 kì vọng lớn).

Việc chuyển thành và làm cực đại sẽ thoả được mong muốn thứ hai.

1. Độ đo sai số khi phân lớp từ công thức lỗi sai TNR, apperant, position recall, accuracy, sensitivity?[tham khảo](https://towardsdatascience.com/machine-learning-an-error-by-any-other-name-a7760a702c4d)
2. Gom nhóm K-Mean, ‘không có điểm nào thay đổi’ liên quan gì đến cực tiểu lỗi?