Tên: Đặng Vũ Phương Uyên

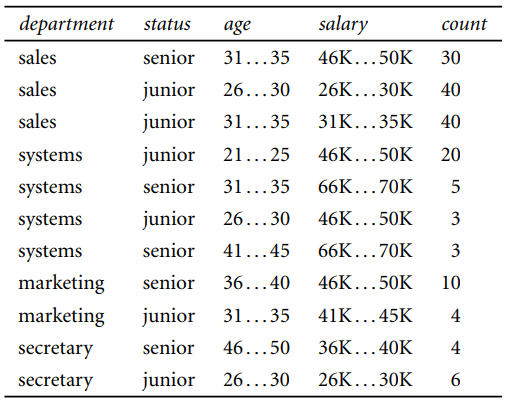
Mssv: 19520345

Bài làm Quizz buổi 6

Bài 1: Tại sao cách phân loại theo trường phái Bayes được gọi là ‘naive’? Hãy trình bày ngắn gọn những ý tưởng chính của bộ phân loại theo naive Bayes.

* Phân loại Naive Bayes được gọi là "naive" vì nó giả định tính độc lập có điều kiện của lớp. Có nghĩa là, ảnh hưởng của một giá trị thuộc tính trên một lớp nhất định là độc lập với các giá trị của các thuộc tính khác. Giả định này được thực hiện để giảm chi phí tính toán, và do đó được coi là “không có”.
* Ý tưởng chính của bộ phân loại theo naive Bayes là cố gắng và phân loại dữ liệu bằng cách tối đa hóa P (X | Ci) P (Ci) (trong đó i là một chỉ số của lớp) bằng cách sử dụng định lý Bayes về xác suất hậu nghiệm.

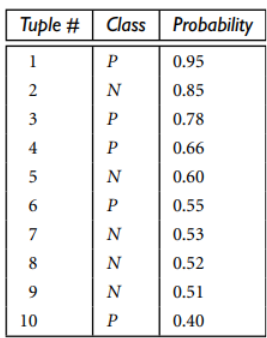
Bài 2: Bảng bên dưới bao gồm dữ liệu huấn luyện từ cơ sở dữ liệu nhân viên. Dữ liệu đã được tổng quát hóa. Ví dụ: ‘31 ... 35’ cho độ tuổi đại diện cho độ tuổi từ 31 đến 35. Đối với một hàng nhất định, số lượng biểu thị số lượng bộ dữ liệu có các giá trị cho phòng ban, trạng thái, tuổi và mức lương được cung cấp trong hàng đó . Cho một bộ dữ liệu có các giá trị ‘system’, ‘26. . . 30, ‘và’ 46–50K ‘cho thược tính department, age, và salary phân loại trạng thái theo kiểu Bayes ban đầu sẽ như thế nào?



Vì khi xét age với giá trị là ‘26 ... 30’ thì status chỉ có một giá trị junior nên ta dùng công thức Laplace để tránh trường hợp xuất hiện 0.

* P(status= ‘seninor’) \* P(X| status = ‘seninor’) = P(status= ‘seninor’) \* P(department= ‘system’| status= ‘seninor’) \* P(age= ‘26…30’| status= ‘seninor’) \* P(salary= ‘46-50K’| status= ‘seninor’)
* P(status= ‘seninor’) = (5+1)/(11+2) = 6/13
* P(department= ‘system’| status= ‘seninor’) = (2+1)/(5+4) = 3/9
* P(age= ‘26…30’| status= ‘seninor’) = (0+1)/(5+6) = 1/11
* P(salary= ‘46-50K’| status= ‘seninor’) = (2+1)/(5+6) = 3/11
* P(status= ‘seninor’) \* P(X| status = ‘seninor’) = 0.0038
* P(status= ‘junior’) \* P(X| status= ‘junior’) = P(status= ‘junior’) \* P(department= ‘system’| status= ‘junior’) \* P(age= ‘26…30’| status= ‘junior’) \* P(salary= ‘46-50K’| status= ‘junior’)
* P(status= ‘junior’) = (6+1)/(11+2) = 7/13
* P(department= ‘system’| status= ‘junior’) = (2+1)/(6+4) = 3/10
* P(age= ‘26…30’| status= ‘junior’) = (3+1)/(6+6) = 4/12
* P(salary= ‘46-50K’| status= ‘junior’) = (2+1)/(6+6) = 3/12
* P(status= ‘junior’) \* P(X| status= ‘junior’) = 0.01346
* Mẫu X(department= ‘system’, age= ‘26…30’, salary= ‘46-50K’) thuộc phân lớp status = ‘junior’

Bài 3: Các bộ dữ liệu của như hình được sắp xếp theo giá trị xác suất giảm dần, do bộ phân loại trả về. Đối với mỗi bộ, tính toán các giá trị cho số lần True Positive (TP), False Positive (FP), True Negative (TN) và False Negative (FN). Tính tỷ lệ dương tính thực (TPR) và tỷ lệ dương tính giả (FPR). Vẽ đường cong ROC cho dữ liệu.



Graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, line chart

Description automatically generated