Nhập môn Học máy và Khai phá dữ liệu
(IT3190)

Nguyễn Nhật Quang

quang.nguyennhat@hust.edu.vn

Trường Đại học Bách Khoa Hà Nội

Viện Công nghệ thông tin và truyền thông

Năm học 2020-2021

Nội dung môn học:

- Giới thiệu về Học máy và Khai phá dữ liệu
- Tiền xử lý dữ liệu
- Đánh giá hiệu năng của hệ thống
- Hồi quy
- **Phân lớp**
  - **Các phương pháp học dựa trên xác suất (Probabilistic learning)**
- Phân cụm
- Phát hiện luật kết hợp

Các phương pháp học dựa trên xác suất

- Các phương pháp thống kê cho bài toán phân loại
- Phân loại dựa trên một mô hình xác suất cơ sở
- Việc phân loại dựa trên khả năng xảy ra (probabilities) của các phân lớp
- Các chủ đề chính:
  - Giới thiệu về xác suất
  - Định lý Bayes
  - Xác suất hậu nghiệm cực đại (Maximum a posteriori)
  - Đánh giá khả năng có thể nhất (Maximum likelihood estimation)
  - Phân loại Naïve Bayes

Các khái niệm cơ bản về xác suất

- Giả sử chúng ta có một thí nghiệm (ví dụ: đổ một quân xúc sắc) mà kết quả của nó mang tính ngẫu nhiên (phụ thuộc vào khả năng có thể xảy ra)
- *Không gian các khả năng* `S`. Tập hợp tất cả các kết quả có thể xảy ra

  Ví dụ: `S= {1,2,3,4,5,6}` đối với thí nghiệm đổ quân xúc sắc
- *Sự kiện* `E`. Một tập con của không gian các khả năng

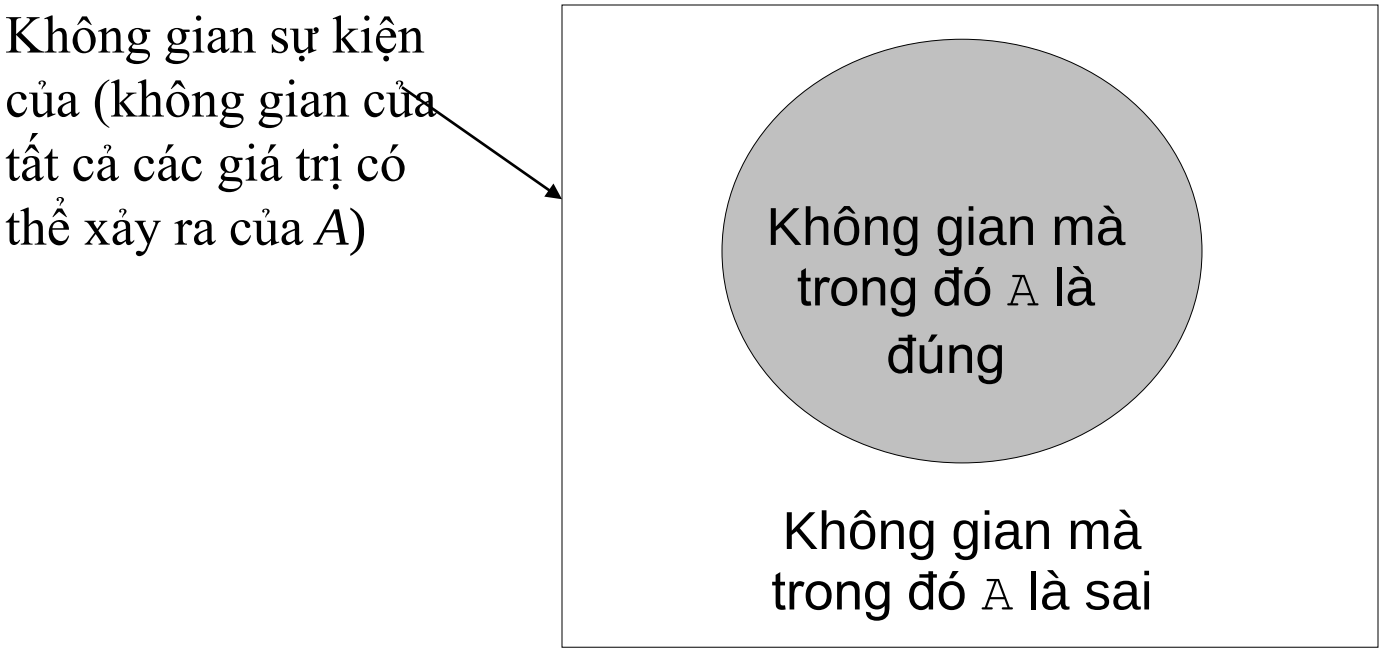
  Ví dụ: `E= {1}`: kết quả quân súc xắc đổ ra là 1

  Ví dụ: `E= {1,3,5}`: kết quả quân súc xắc đổ ra là một số lẻ
- *Không gian các sự kiện* `W`. Không gian (thế giới) mà các kết quả của sự kiện có thể xảy ra

  Ví dụ: `W` bao gồm tất cả các lần đổ súc xắc
- *Biến ngẫu nhiên* `A`. Một biến ngẫu nhiên biểu diễn (diễn đạt) một sự kiện, và có một mức độ về khả năng xảy ra sự kiện này

Biểu diễn xác suất

P(A): "Phần của không gian (thế giới) mà trong đó A là đúng"



[http://www.cs.cmu.edu/~awm/tutorials]

Các biến ngẫu nhiên 2 giá trị

- Một biến ngẫu nhiên 2 giá trị (nhị phân) có thể nhận một trong 2 giá trị đúng (`true`) hoặc sai (`false`)
- Các tiên đề
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$
  - $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- Các hệ quả
  - $P(not\ A) \equiv P(\sim A) = 1 - P(A)$
  - $P(A) = P(A \wedge B) + P(A \wedge \sim B)$

Các biến ngẫu nhiên đa trị

Một biến ngẫu nhiên nhiều giá trị có thể nhận một trong số k (>2) giá trị $\{v_1, v_2, \ldots, v_k\}$

$$P(A = v_i \wedge A = v_j) = 0 \text{ if } i \neq j$$

$$P(A=v_1 \vee A=v_2 \vee \ldots \vee A=v_k) = 1$$

$$P(A = v_1 \vee A = v_2 \vee \ldots \vee A = v_i) = \sum_{j=1}^{i} P(A = v_j)$$

$$\sum_{j=1}^{k} P(A = v_j) = 1$$

$$P\left(B \wedge \left[A = v_1 \vee A = v_2 \vee \ldots \vee A = v_i\right]\right) = \sum_{j=1}^{i} P(B \wedge A = v_j)$$

[http://www.cs.cmu.edu/~awm/tutorials]

Xác suất có điều kiện (1)

- `P(A|B)` là phần của không gian (thế giới) mà trong đó `A` là đúng, với điều kiện (đã biết) là `B` đúng
- Ví dụ
  - `A`: Tôi sẽ đi đá bóng vào ngày mai
  - `B`: Trời sẽ không mưa vào ngày mai
  - `P(A|B)`: Xác suất của việc tôi sẽ đi đá bóng vào ngày mai nếu (đã biết rằng) trời sẽ không mưa (vào ngày mai)

Xác suất có điều kiện (2)

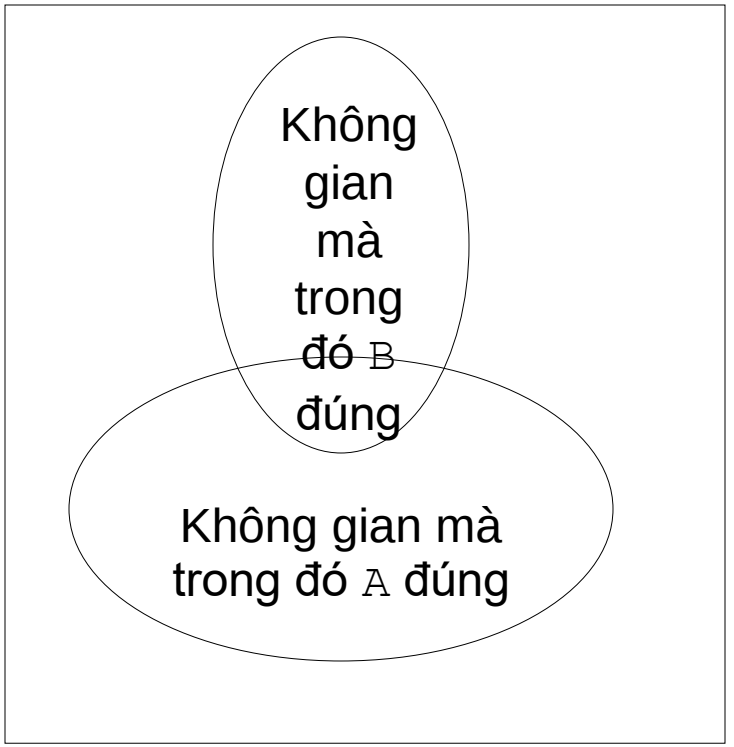
Định nghĩa: $P(A|B) = \frac{P(A,B)}{P(B)}$

Các hệ quả:

$P(A,B) = P(A|B).P(B)$

$P(A|B) + P(\sim A|B) = 1$

$$\sum_{i=1}^{k} P(A = v_i | B) = 1$$



Các biến độc lập về xác suất (1)

- Hai sự kiện A và B được gọi là ***độc lập về xác suất*** nếu xác suất của sự kiện A là như nhau đối với các trường hợp:
  - Khi sự kiện B xảy ra, hoặc
  - Khi sự kiện B không xảy ra, hoặc
  - Không có thông tin (không biết gì) về việc xảy ra của sự kiện B
- Ví dụ
  - A: Tôi sẽ đi đá bóng vào ngày mai
  - B: Tuấn sẽ tham gia trận đá bóng ngày mai
  - P(A|B) = P(A)

  → "Dù Tuấn có tham gia trận đá bóng ngày mai hay không cũng không ảnh hưởng tới quyết định của tôi về việc đi đá bóng ngày mai."

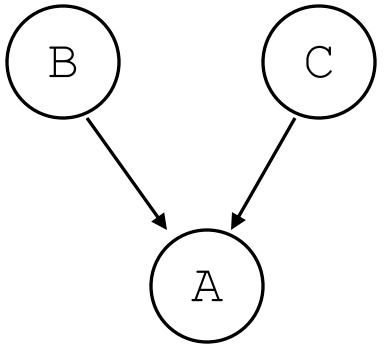
Các biến độc lập về xác suất (2)

Từ định nghĩa của các biến độc lập về xác suất $P(A|B)=P(A)$, chúng ta thu được các luật như sau

- $P(\neg A|B) = P(\neg A)$
- $P(B|A) = P(B)$
- $P(A,B) = P(A).\ P(B)$
- $P(\neg A,B) = P(\neg A).\ P(B)$
- $P(A,\neg B) = P(A).\ P(\neg B)$
- $P(\neg A,\neg B) = P(\neg A).\ P(\neg B)$

Xác suất có điều kiện với >2 biến

- `P(A|B,C)` là xác suất của `A` đối với (đã biết) `B` và `C`
- Ví dụ
  - `A`: Tôi sẽ đi dạo bờ sông vào sáng mai
  - `B`: Thời tiết sáng mai rất đẹp
  - `C`: Tôi sẽ dậy sớm vào sáng mai
  - `P(A|B,C)`: Xác suất của việc tôi sẽ đi dạo dọc bờ sông vào sáng mai, nếu (đã biết rằng) thời tiết sáng mai rất đẹp và tôi sẽ dậy sớm vào sáng mai



`P(A|B,C)`

Độc lập có điều kiện

- Hai biến A và C được gọi là ***độc lập có điều kiện*** đối với biến B, nếu xác suất của A đối với B bằng xác suất của A đối với B và C
- Công thức định nghĩa: $P(A|B,C) = P(A|B)$
- Ví dụ
  - A: Tôi sẽ đi đá bóng vào ngày mai
  - B: Trận đá bóng ngày mai sẽ diễn ra trong nhà
  - C: Ngày mai trời sẽ không mưa
  - $P(A|B,C)=P(A|B)$

    → Nếu biết rằng trận đấu ngày mai sẽ diễn ra trong nhà, thì xác suất của việc tôi sẽ đi đá bóng ngày mai không phụ thuộc vào thời tiết

Các quy tắc quan trọng của xác suất

- Quy tắc chuỗi (chain rule)
  - $P(A,B) = P(A|B).P(B) = P(B|A).P(A)$
  - $P(A|B) = P(A,B)/P(B) = P(B|A).P(A)/P(B)$
  - $P(A,B|C) = P(A,B,C)/P(C) = P(A|B,C).P(B,C)/P(C)$

    $= P(A|B,C).P(B|C)$
- Độc lập về xác suất và độc lập có điều kiện
  - $P(A|B) = P(A)$; nếu A và B là độc lập về xác suất
  - $P(A,B|C) = P(A|C).P(B|C)$; nếu A và B là độc lập có điều kiện đối với C
  - $P(A_1,\ldots,A_n|C) = P(A_1|C)\ldots P(A_n|C)$; nếu $A_1,\ldots,A_n$ là độc lập có điều kiện đối với C

Định lý Bayes

$$P(h|D) = \frac{P(D|h).P(h)}{P(D)}$$

- `P(h)`: Xác suất trước (tiên nghiệm) của giả thiết (phân loại) `h`
- `P(D)`: Xác suất trước (tiên nghiệm) của việc quan sát được dữ liệu `D`
- `P(D|h)`: Xác suất (có điều kiện) của việc quan sát được dữ liệu `D`, nếu biết giả thiết (phân loại) `h` là đúng
- `P(h|D)`: Xác suất (có điều kiện) của giả thiết (phân loại) `h` là đúng, nếu quan sát được dữ liệu `D`
  - **Các phương pháp phân loại dựa trên xác suất sẽ sử dụng xác suất có điều kiện (*posterior probability*) này!**

Định lý Bayes – Ví dụ (1)

Giả sử chúng ta có tập dữ liệu sau (dự đoán 1 người có chơi tennis)?

Ngày	Ngoài trời	Nhiệt độ	Độ ẩm	Gió	Chơi tennis
N1	Nắng	Nóng	Cao	Yếu	Không
N2	Nắng	Nóng	Cao	Mạnh	Không
N3	Âm u	Nóng	Cao	Yếu	Có
N4	Mưa	Bình thường	Cao	Yếu	Có
N5	Mưa	Mát mẻ	Bình thường	Yếu	Có
N6	Mưa	Mát mẻ	Bình thường	Mạnh	Không
N7	Âm u	Mát mẻ	Bình thường	Mạnh	Có
N8	Nắng	Bình thường	Cao	Yếu	Không
N9	Nắng	Mát mẻ	Bình thường	Yếu	Có
N10	Mưa	Bình thường	Bình thường	Yếu	Có
N11	Nắng	Bình thường	Bình thường	Mạnh	Có
N12	Âm u	Bình thường	Cao	Mạnh	Có

[Mitchell, 1997]

Định lý Bayes – Ví dụ (2)

- Dữ liệu `D`. *Ngoài trời* là *nắng* và *Gió* là *mạnh*
- Giả thiết (phân loại) `h`. Anh ta chơi tennis
- Xác suất trước `P(h)`. Xác suất rằng anh ta chơi tennis (bất kể *Ngoài trời* như thế nào và *Gió* ra sao)
- Xác suất trước `P(D)`. Xác suất rằng *Ngoài trời* là *nắng* và *Gió* là *mạnh*
- `P(D|h)`. Xác suất *Ngoài trời* là *nắng* và *Gió* là *mạnh*, nếu biết rằng anh ta chơi tennis
- `P(h|D)`. Xác suất anh ta chơi tennis, nếu biết rằng *Ngoài trời* là *nắng* và *Gió* là *mạnh*

  → Chúng ta quan tâm đến giá trị xác suất sau (*posterior probability*) này!

Xác suất hậu nghiệm cựu đại (MAP)

- Với một tập các giả thiết (các phân lớp) có thể `H`, hệ thống học sẽ tìm ***giả thiết có thể xảy ra nhất (the most probable hypothesis)*** `h(∈H)` đối với các dữ liệu quan sát được `D`
- Giả thiết `h` này được gọi là giả thiết có xác suất hậu nghiệm cực đại (**Maximum a posteriori – MAP**)

$$h_{MAP} = \arg\max_{h \in H} P(h \mid D)$$

$$h_{MAP} = \arg\max_{h \in H} \frac{P(D \mid h).P(h)}{P(D)}$$ (bởi định lý Bayes)

$$h_{MAP} = \arg\max_{h \in H} P(D \mid h).P(h)$$ (`P(D)` là như nhau đối với các giả thiết `h`)

MAP – Ví dụ

- Tập H bao gồm 2 giả thiết (có thể)
  - h_1: Anh ta chơi tennis
  - h_2: Anh ta không chơi tennis
- Tính giá trị của 2 xác xuất có điều kiện: $P(h_1|D)$, $P(h_2|D)$
- Giả thiết có thể nhất $h_{MAP}=h_1$ nếu $P(h_1|D) \geq P(h_2|D)$; ngược lại thì $h_{MAP}=h_2$
- Bởi vì $P(D)=P(D,h_1)+P(D,h_2)$ là như nhau đối với cả 2 giả thiết h_1 và h_2, nên có thể bỏ qua đại lượng $P(D)$
- Vì vậy, cần tính 2 biểu thức: $P(D|h_1).P(h_1)$ và $P(D|h_2).P(h_2)$, và đưa ra quyết định tương ứng
  - Nếu $P(D|h_1).P(h_1) \geq P(D|h_2).P(h_2)$, thì kết luận là anh ta chơi tennis
  - Ngược lại, thì kết luận là anh ta không chơi tennis

Đánh giá khả năng có thể nhất (MLE)

- Phương pháp MAP: Với một tập các giả thiết có thể H, cần tìm một giả thiết cực đại hóa giá trị: $P(D|h).P(h)$
- Giả sử (assumption) trong phương pháp **đánh giá khả năng có thể nhất (Maximum likelihood estimation – MLE)**: Tất cả các giả thiết đều có giá trị xác suất trước như nhau: $P(h_i)=P(h_j)$, $\forall h_i,h_j \in H$
- Phương pháp MLE tìm giả thiết cực đại hóa giá trị $P(D|h)$; trong đó $P(D|h)$ được gọi là *khả năng có thể (likelihood)* của dữ liệu D đối với h
- Giả thiết có khả năng nhất (maximum likelihood hypothesis)

$$h_{ML} = \arg\max_{h \in H} P(D|h)$$

MLE – Ví dụ

- Tập H bao gồm 2 giả thiết có thể
  - h_1: Anh ta chơi tennis
  - h_2: Anh ta không chơi tennis

  D: Tập dữ liệu (các ngày) mà trong đó thuộc tính *Ngoài trời* có giá trị *Nắng* và thuộc tính *Gió* có giá trị *Mạnh*
- Tính 2 giá trị khả năng xảy ra (likelihood values) của dữ liệu D đối với 2 giả thiết: $P(D|h_1)$ và $P(D|h_2)$
  - $P(\text{Ngoài trời}=\text{Nắng}, \text{Gió}=\text{Mạnh}|h_1) = 1/8$
  - $P(\text{Ngoài trời}=\text{Nắng}, \text{Gió}=\text{Mạnh}|h_2) = 1/4$
- Giả thiết MLE $h_{MLE}=h_1$ nếu $P(D|h_1) \geq P(D|h_2)$; và ngược lại thì $h_{MLE}=h_2$

  → Bởi vì $P(\text{Outlook}=\text{Sunny}, \text{Wind}=\text{Strong}|h_1) < P(\text{Outlook}=\text{Sunny}, \text{Wind}=\text{Strong}|h_2)$, hệ thống kết luận rằng: *Anh ta sẽ không chơi tennis!*

Phân loại Naïve Bayes (1)

- Biểu diễn bài toán phân loại (classification problem)
  - Một tập học `D_train`, trong đó mỗi ví dụ học `x` được biểu diễn là một vectơ `n` chiều: $(x_1, x_2, ..., x_n)$
  - Một tập xác định các nhãn lớp: $C=\{c_1, c_2, ..., c_m\}$
  - Với một ví dụ (mới) `z`, thì `z` sẽ được phân vào lớp nào?
- Mục tiêu: Xác định phân lớp có thể (phù hợp) nhất đối với `z`

$$c_{MAP} = \arg\max_{c_i \in C} P(c_i \mid z)$$

$$c_{MAP} = \arg\max_{c_i \in C} P(c_i \mid z_1, z_2, ..., z_n)$$

$$c_{MAP} = \arg\max_{c_i \in C} \frac{P(z_1, z_2, ..., z_n \mid c_i).P(c_i)}{P(z_1, z_2, ..., z_n)}$$ (bởi định lý Bayes)

Phân loại Naïve Bayes (2)

- Để tìm được phân lớp có thể nhất đối với z…

$$c_{MAP} = \arg\max_{c_i \in C} P(z_1, z_2, ..., z_n \mid c_i).P(c_i)$$

($P(z_1, z_2, ..., z_n)$ là như nhau với các lớp)

- **Giả sử (assumption) trong phương pháp phân loại Naïve Bayes**. Các thuộc tính là *độc lập có điều kiện (conditionally independent)* đối với các lớp

$$P(z_1, z_2, ..., z_n \mid c_i) = \prod_{j=1}^{n} P(z_j \mid c_i)$$

- Phân loại Naïve Bayes tìm phân lớp có thể nhất đối với z

$$c_{NB} = \arg\max_{c_i \in C} P(c_i).\prod_{j=1}^{n} P(z_j \mid c_i)$$

Phân loại Naïve Bayes – Giải thuật

- Giai đoạn học (training phase), sử dụng một tập học

  Đối với mỗi phân lớp có thể (mỗi nhãn lớp) $c_i \in C$

  - Tính giá trị xác suất trước: $P(c_i)$
  - Đối với mỗi giá trị thuộc tính x_j, tính giá trị xác suất xảy ra của giá trị thuộc tính đó đối với một phân lớp c_i: $P(x_j \mid c_i)$

- Giai đoạn phân lớp (classification phase), đối với một ví dụ mới
  - Đối với mỗi phân lớp $c_i \in C$, tính giá trị của biểu thức:

$$P(c_i).\prod_{j=1}^{n} P(x_j \mid c_i)$$

  - Xác định phân lớp của z là lớp có thể nhất c^*

$$c^* = \arg\max_{c_i \in C} P(c_i).\prod_{j=1}^{n} P(x_j \mid c_i)$$

Phân lớp Naïve Bayes – Ví dụ (1)

Một sinh viên trẻ với thu nhập trung bình và mức đánh giá tín dụng bình thường sẽ mua một cái máy tính?

Rec. ID	Age	Income	Student	Credit_Rating	Buy_Computer
1	Young	High	No	Fair	No
2	Young	High	No	Excellent	No
3	Medium	High	No	Fair	Yes
4	Old	Medium	No	Fair	Yes
5	Old	Low	Yes	Fair	Yes
6	Old	Low	Yes	Excellent	No
7	Medium	Low	Yes	Excellent	Yes
8	Young	Medium	No	Fair	No
9	Young	Low	Yes	Fair	Yes
10	Old	Medium	Yes	Fair	Yes
11	Young	Medium	Yes	Excellent	Yes
12	Medium	Medium	No	Excellent	Yes
13	Medium	High	Yes	Fair	Yes
14	Old	Medium	No	Excellent	No

Phân lớp Naïve Bayes – Ví dụ (2)

- Biểu diễn bài toán phân loại
  - z = (Age=Young, Income=Medium, Student=Yes, Credit_Rating=Fair)
  - Có 2 phân lớp có thể: c_1 ("Mua máy tính") và c_2 ("Không mua máy tính")
- Tính giá trị xác suất trước cho mỗi phân lớp
  - P(c_1) = 9/14
  - P(c_2) = 5/14
- Tính giá trị xác suất của mỗi giá trị thuộc tính đối với mỗi phân lớp
  - P(Age=Young|c_1) = 2/9; P(Age=Young|c_2) = 3/5
  - P(Income=Medium|c_1) = 4/9; P(Income=Medium|c_2) = 2/5
  - P(Student=Yes|c_1) = 6/9; P(Student=Yes|c_2) = 1/5
  - P(Credit_Rating=Fair|c_1) = 6/9; P(Credit_Rating=Fair|c_2) = 2/5

Phân lớp Naïve Bayes – Ví dụ (3)

- Tính toán xác suất có thể xảy ra (likelihood) của ví dụ z đối với mỗi phân lớp
  - Đối với phân lớp c_1

    $P(z|c_1) = P(\text{Age=Young}|c_1).P(\text{Income=Medium}|c_1).P(\text{Student=Yes}|c_1).$
    $P(\text{Credit\_Rating=Fair}|c_1) = (2/9).(4/9).(6/9).(6/9) = 0.044$

  - Đối với phân lớp c_2

    $P(z|c_2) = P(\text{Age=Young}|c_2).P(\text{Income=Medium}|c_2).P(\text{Student=Yes}|c_2).$
    $P(\text{Credit\_Rating=Fair}|c_2) = (3/5).(2/5).(1/5).(2/5) = 0.019$

- Xác định phân lớp có thể nhất (the most probable class)
  - Đối với phân lớp c_1

    $P(c_1).P(z|c_1) = (9/14).(0.044) = 0.028$

  - Đối với phân lớp c_2

    $P(c_2).P(z|c_2) = (5/14).(0.019) = 0.007$

  → Kết luận: *Anh ta (z) sẽ mua một máy tính*!

Phân lớp Naïve Bayes – Vấn đề (1)

- Nếu không có ví dụ nào gắn với phân lớp c_i có giá trị thuộc tính x_j…

  $P(x_j \mid c_i)=0$, và vì vậy: $$P(c_i).\prod_{j=1}^{n} P(x_j \mid c_i) = 0$$

- Giải pháp: Sử dụng phương pháp Bayes để ước lượng $P(x_j \mid c_i)$

$$P(x_j \mid c_i) = \frac{n(c_i, x_j) + mp}{n(c_i) + m}$$

  - $n(c_i)$: số lượng các ví dụ học gắn với phân lớp c_i
  - $n(c_i, x_j)$: số lượng các ví dụ học gắn với phân lớp c_i có giá trị thuộc tính x_j
  - p: ước lượng đối với giá trị xác suất $P(x_j \mid c_i)$

    → Các ước lượng đồng mức: $p=1/k$, với thuộc tính f_j có k giá trị có thể
  - m: một hệ số (trọng số)

    → Để bổ sung cho $n(c_i)$ các ví dụ thực sự được quan sát với thêm m mẫu ví dụ với ước lượng p

Phân lớp Naïve Bayes – Vấn đề (2)

- Giới hạn về độ chính xác trong tính toán của máy tính
  - $P(x_j \mid c_i) < 1$, đối với mọi giá trị thuộc tính x_j và phân lớp c_i
  - Vì vậy, khi số lượng các giá trị thuộc tính là rất lớn, thì:

$$\lim_{n\to\infty}\left(\prod_{j=1}^{n} P(x_j \mid c_i)\right) = 0$$

- Giải pháp: Sử dụng hàm lôgarit cho các giá trị xác suất

$$c_{NB} = \arg\max_{c_i \in C}\left(\log\left[P(c_i).\prod_{j=1}^{n} P(x_j \mid c_i)\right]\right)$$

$$c_{NB} = \arg\max_{c_i \in C}\left(\log P(c_i) + \sum_{j=1}^{n} \log P(x_j \mid c_i)\right)$$

Phân loại văn bản bằng NB (1)

- Biểu diễn bài toán phân loại văn bản
  - Tập học D_train, trong đó mỗi ví dụ học là một biểu diễn văn bản gắn với một nhãn lớp: D_train = {(d_k, c_i)}
  - Một tập các nhãn lớp xác định: C = {c_i}
- Giai đoạn học
  - Từ tập các văn bản trong D_train, trích ra tập các từ khóa (keywords/terms): T = {t_j}
  - Gọi D_c_i ($\subseteq$D_train) là tập các văn bản trong D_train có nhãn lớp c_i
  - Đối với mỗi phân lớp c_i
    - Tính giá trị xác suất trước của phân lớp c_i: $P(c_i) = \frac{|D\_c_i|}{|D|}$
    - Đối với mỗi từ khóa t_j, tính xác suất từ khóa t_j xuất hiện đối với lớp c_i

$$P(t_j \mid c_i) = \frac{\left(\sum_{d_k \in D\_c_i} n(d_k, t_j)\right) + 1}{\left(\sum_{d_k \in D\_c_i} \sum_{t_m \in T} n(d_k, t_m)\right) + |T|}$$

($n(d_k, t_j)$: số lần xuất hiện của từ khóa t_j trong văn bản d_k)

Phân loại văn bản bằng NB (2)

- Giai đoạn phân lớp đối với một văn bản mới d
  - Từ văn bản d, trích ra tập T_d gồm các từ khóa (keywords) t_j đã được định nghĩa trong tập T ($T\_d \subseteq T$)
  - **Giả sử (assumption)**. Xác suất từ khóa t_j xuất hiện đối với lớp c_i là độc lập đối với vị trí của từ khóa đó trong văn bản

    P(t_j ở vị trí $k|c_i$) = P(t_j ở vị trí $m|c_i$), $\forall k,m$
  - Đối với mỗi phân lớp c_i, tính giá trị likelihood của văn bản d đối với c_i

$$P(c_i).\prod_{t_j \in T\_d} P(t_j \mid c_i)$$

  - Phân lớp văn bản d thuộc vào lớp c^*

$$c^* = \arg\max_{c_i \in C} P(c_i).\prod_{t_j \in T\_d} P(t_j \mid c_i)$$

Phân loại Naïve Bayes – Tổng kết

- Một trong các phương pháp học máy được áp dụng phổ biến nhất trong thực tế
- Dựa trên định lý Bayes
- Việc phân loại dựa trên các giá trị xác suất của các khả năng xảy ra của các giả thiết (phân loại)
- Mặc dù đặt giả sử về sự độc lập có điều kiện của các thuộc tính đối với các phân lớp, nhưng phương pháp phân loại Naïve Bayes vẫn thu được các kết quả phân loại tốt trong nhiều lĩnh vực ứng dụng thực tế
- Khi nào nên sử dụng?
  - Có một tập huấn luyện có kích thước lớn hoặc vừa
  - Các ví dụ được biểu diễn bởi một số lượng lớn các thuộc tính
  - **Các thuộc tính độc lập có điều kiện đối với các phân lớp**