MACHINE LEARNING

1211635

MAHASARAKHAM UNIVERSITY

BAYESIAN LEARNING

Naive Bayes

Olarik Surinta, PhD.
Lecturer

Bayesian Learning

- การเรียนรู้แบบเบย์ (Bayesian Learning) เป็นการจำแนกประเภท (Classification) รูปแบบหนึ่งที่อาศัยหลักการของความน่าจะเป็น (Probability) เข้ามาช่วยในการหาคำตอบของประเภทตัวอย่างใหม่
- การเรียนรู้แบบเบย์เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด

Bayes' theorem

- ทฤษฎีของเบย์ (Bayes' theorem) ถูกนำเสนอโดย Thomas Bayes

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

- $P(h)$ = prior probability of hypothesis h
- $P(D)$ = prior probability of training data D
- $P(h|D)$ = probability of h given D
- $P(D|h)$ = probability of D given h

Bayes' theorem

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

- $P(h)$ = prior probability of hypothesis h
- $P(D)$ = prior probability of training data D
- $P(h|D)$ = probability of h given D
- $P(D|h)$ = probability of D given h

โดย

D	แทนข้อมูลที่นำมาใช้ในการคำนวณการแจกแจงความน่าจะเป็น posteriori probability ของสมมติฐาน h คือ P(h|D) ตามทฤษฎี
P(h)	คือความน่าจะเป็นก่อนหน้าของสมมติฐาน h
P(D)	คือความน่าจะเป็นก่อนหน้าของชุดข้อมูลตัวอย่าง D
P(h|D)	คือความน่าจะเป็นของ h เมื่อรู้ D
P(D|h)	คือความน่าจะเป็นของ D เมื่อรู้ h

Bayes' theorem

Likelihood

Prior probability

Posterior probability

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

Predictor prior probability

$$\boldsymbol{P(h|D)} = P(D_1|h) \times P(D_2|h) \times \ldots \times P(D_n|h) \times P(h)$$

Basic formulas for probabilities

- *Product Rule*: probability P(A ^ B) of a conjunction of two events A and B:

$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

- *Sum Rule*: probability of disjunction of two events A and B:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

- *Theorem of total probability*: if events $A_1, \ldots, A_n$ are mutually exclusive with $\Sigma_{i=1}^{n} P(A_i) = 1$, then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Brute force MAP hypothesis learner

1) For each hypothesis h in H, calculate the posterior probability

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

2) Output the hypothesis h_{MAP} with the highest posterior probability

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}} P(h|D)$$

Bayes’ theorem

- ตัวอย่างการคำนวณเพื่อเลือกสมมติฐานโดยกฎของเบย์
  - คนไข้ไปโรงพยาบาลเพื่อตรวจหา “โรคมะเร็ง (cancer)” โดยผลการตรวจจากห้องตรวจมีผลเป็นบวก (+) ซึ่งให้ความถูกต้อง 98% ในกรณีที่มีโรคนั้นอยู่จริง และมีผลเป็นลบ (-) ซึ่งให้ความถูกต้อง 97% ในกรณีที่ไม่มีโรคนั้น โดยมีจำนวนผู้ที่เป็นโรคมะเร็ง 0.008 ของประชากรทั้งหมด อยากทราบว่า คนไข้มีโอกาสเป็นโรคมะเร็งหรือไม่

Bayes' theorem

P(cancer) = 0.008

P(~cancer) = 0.992

P(+|cancer) = 0.98

P(-|cancer) = 0.02

P(+|~cancer) = 0.03

P(-|~cancer) = 0.97

ดังนั้น เราสามารถคำนวณค่าความน่าจะเป็นของสมมติฐานว่าคนไข้เป็นโรคมะเร็ง หรือไม่ได้เป็นโรคมะเร็ง เมื่อทราบผลตรวจเป็นบวก (+) โดยใช้กฎของเบย์ ดังนี้

Basic formulas for probabilities

- *Product Rule*: probability P(A ^ B) of a conjunction of two events A and B:

$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

- *Sum Rule*: probability of disjunction of two events A and B:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

- *Theorem of total probability*: if events $A_1, \ldots, A_n$ are mutually exclusive with $\Sigma_{i=1}^{n} P(A_i) = 1$, then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Brute force MAP hypothesis learner

1) For each hypothesis h in H, calculate the posterior probability

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

2) Output the hypothesis h_{MAP} with the highest posterior probability

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}}\, P(h|D)$$

Bayes' theorem

- **<u>สมมติฐานที่ 1</u>** คนไข้**เป็น**โรคมะเร็งจริง เมื่อมีผลการตรวจเป็นบวก

เขียนแทนด้วย P(cancer | +)

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

$$P(\text{cancer} \mid +) = \frac{P(+ \mid \text{cancer})\, P(\text{cancer})}{P(+)}$$

$$= 0.98 * 0.008$$

$$= \mathbf{0.0078}$$

Bayes' theorem

- **<u>สมมติฐานที่ 2</u>** คนไข้**ไม่เป็น**โรคมะเร็งจริง เมื่อมีผลการตรวจเป็นบวก

เขียนแทนด้วย P(~cancer | +)

$$P(\sim cancer \mid +) = \frac{P(+ \mid \sim cancer)\, P(\sim cancer)}{P(+)}$$

$$= 0.03 * 0.992$$

$$= \mathbf{0.0298}$$

Bayes' theorem

- จากกรณีนี้มีเพียง 2 สมมติฐาน นั่นแสดงว่าผลรวมของความน่าจะเป็นระหว่าง P(cancer|+) และ P(~cancer|+) จะมีค่าเท่ากับ **1**
- **ดังนั้น** สามารถ Normalize ค่าของ P(cancer|+) และ P(~cancer|+) ดังนี้
- P(cancer|+) = 0.0078/(0.0078+0.0298) = **0.21**
- P(~cancer|+) = 0.0298/(0.0078+0.0298) = **0.79**

Bayes' theorem

- จากการคำนวณสรุปได้ดังนี้
  - สมมติฐานที่ 1 มีค่าความน่าจะเป็นเท่ากับ 0.21
  - สมมติฐานที่ 2 มีค่าความน่าจะเป็นเท่ากับ 0.79
- เลือกตอบสมมติฐานที่ 2 เนื่องจากมีค่ามากกว่า
- **ดังนั้น** สมมติฐานที่ว่าคนไข้ไม่เป็นโรคมะเร็งเมื่อทราบผลตรวจที่เป็นบวก ด้วยความน่าจะเป็น 0.79 จึงถูกเลือกนำมาเป็นคำตอบ

Naive Bayes

ความน่าจะเป็นที่ B เกิดก่อนและ A เกิดตามมา

ความน่าจะเป็นที่ A และ B เกิดร่วมกัน

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

P(A|B) คือ ค่า conditional probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา

P(A ∩ B) คือ ค่า joint probability หรือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

P(B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น

Cr. http://dataminingtrend.com/2014/naive-bayes/

Naive Bayes

- ในลักษณะเดียวกันเราจะเขียน P(B|A) หรือค่าความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้นก่อนและเหตุการณ์ B เกิดขึ้นตามมาที่หลังได้เป็น

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Naive Bayes

- จากทั้ง 2 แบบจะเห็นว่ามีค่า P(A ∩ B) ที่เหมือนกันอยู่ดังนั้นเราสามารถเขียนสมการของ P(A ∩ B) ได้เป็นดังนี้

$$P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$$

$$\boxed{P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}}$$

Bayes Theorem

No	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	mild	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Naive Bayes

P(**play** = yes) = 9/14 = 0.64
P(**play** = no) = 5/14 = 0.36

attribute	**play** = yes	**play** = no
outlook = sunny	2/9 = 0.22	3/5 = 0.60
outlook = overcast	4/9 = 0.45	0/5 = 0.00
outlook = rainy	3/9 = 0.33	2/5 = 0.40
temperature = hot	2/9 = 0.22	2/5 = 0.40
temperature = mild	4/9 = 0.45	2/5 = 0.40
temperature = cool	3/9 = 0.33	1/5 = 0.20
humidity = high	3/9 = 0.33	4/5 = 0.80
humidity = normal	6/9 = 0.67	1/5 = 0.20
windy = TRUE	3/9 = 0.33	3/5 = 0.60
windy = FALSE	6/9 = 0.67	2/5 = 0.40

ตารางนี้ก็คือ Model ของอัลกอริทึม Naive Bayes ที่สร้างจาก Training data

Naive Bayes

- ทดสอบโมเดลโดยใช้ข้อมูลชุดทดสอบ (Test data) ในกรณีนี้ ทำการทดสอบโดยใช้ ดังนี้
  - แอตทริบิวต์ outlook = s**unny**
  - แอตทริบิวต์ temperature = **hot**
  - แอตทริบิวต์ humidity = **high**
  - แอตทริบิวต์ windy = **FALSE**
  -

Naive Bayes

- ดังนั้น เราจะต้องคำนวณค่าความน่าจะเป็นที่มีแอตทริบิวต์เหล่านี้แล้วตอบคลาส play=yes

P(**play** = yes|A) = P(**outlook** = sunny|play = yes) x P(**temperature** = hot|play = yes) x
P(**humidity** = high|play = yes) x P(**windy** = FALSE|play = yes) x
P(play = yes)
= 0.22 x 0.22 x 0.33 x 0.67 x 0.64
= 0.0068

- และคำนวณค่าความน่าจะเป็นที่มีแอตทริบิวต์เหล่านี้แล้วตอบคลาส play=no

P(**play** = no|A) = P(**outlook** = sunny|play = no) x P(**temperature** = hot|play = no) x
P(**humidity** = high|play = no) x P(**windy** = FALSE|play = no) x
P(play = no)
= 0.60 x 0.40 x 0.80 x 0.40 x 0.36
= 0.0276

Naive Bayes

- เมื่อเปรียบเทียบค่าความน่าจะเป็นที่ได้จาก 2 คลาส แล้วพบว่าค่า
  - p(play=no|A) = 0.0276
  - p(play=yes|A) = 0.0068
- ดังนั้นคำตอบที่ได้จากการใช้ Model นี้คือ **play=no** เนื่องจากมีค่าความน่าจะเป็นสูงกว่า

References

- http://mis.csit.sci.tsu.ac.th/noppamas/download/DataMining/DataMiningCh7V1.pdf
- http://dataminingtrend.com/2014/naive-bayes/
- https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/
- http://www.cs.cmu.edu/afs/cs/project/theo-20/www/mlbook/ch6.pdf
-