

เนื้อหาการบรรยายครั้งที่ 8

- การเรียนรู้ด้วยเครื่อง (Machine Learning)

การเรียนรู้ด้วยเครื่อง

การเรียนรู้ของเครื่อง

ประเภทของการเรียนรู้

Naïve bayes classifier

Decision tree

Nearest neighbor classification

การเรียนรู้ของเครื่อง

- การพยายามทำให้เครื่องคอมพิวเตอร์หรือเครื่องอิเล็กทรอนิกส์ให้สามารถทำงานที่มีความฉลาดได้
- รสร้างความสามารถในการเรียนรู้จากข้อมูลเข้า
- เป็นการเรียนรู้แบบ inductive learning
 - เรียนรู้จากข้อมูลบางส่วน เพื่อนำไปใช้หาความจริงที่เป็นสากล

การเรียนรู้ของเครื่อง

มนุษย์สามารถเรียนรู้ได้จาก

- คำสอนของครูหรือผู้รู้ (Supervised Learning)
- ประสบการณ์ในอดีต (Learning from experience)
- การจดจำ (Rote learning)
- การรับคำสั่งและการฝึกฝน (Learning by doing)

สถาปัตยกรรมการเรียนรู้ของเครื่อง

- สร้างขั้นตอนการเรียนรู้โดยอาศัยหลักการการเรียนรู้ของมนุษย์
- รับข้อมูลเข้าจำนวนมากพอที่จะทำให้สร้างแนวคิด (concept) หรือรูปแบบ (pattern) ของชุดข้อมูลนั้น
- ชุดข้อมูลที่ใช้สำหรับการเรียนรู้ เรียกว่า ข้อมูลฝึก (training set/data)
- ประเมินคุณภาพของแนวคิดที่ได้ โดยนำข้อมูลมาทดสอบ เรียกว่า ข้อมูลทดสอบ (test set/data)
- ข้อมูลฝึกและข้อมูลทดสอบควรเป็นคนละชุดกัน
- การประเมินประสิทธิภาพในแง่ความถูกต้อง ต้องมีผลเฉลย (labeled data) สำหรับข้อมูลชุดทดสอบ ที่จัดเตรียมไว้

ประเภทของการเรียนรู้

ขั้นตอนวิธีการเรียนรู้

- แบบมีผลเฉลย (Supervised learning)
- แบบไม่มีผลเฉลย (Unsupervised learning)
- แบบกึ่งมีผลเฉลย (Semi-supervised learning)

การเรียนรู้แบบมีผลเฉลย

- เปรียบได้กับการเรียนรู้แบบมีครูฝึกทำหน้าที่ให้คำเฉลย
- ต้องมีการเตรียมข้อมูลฝึกให้มีผลเฉลย
- ขั้นตอนวิธีใช้ข้อมูลฝึกในการค้นหาและสร้างแนวคิดของข้อมูลฝึกแต่ละประเภท
- เรียนรู้ในการจำแนกข้อมูล โดยมีพื้นฐานการพยากรณ์จากข้อมูลทั้งหมดในชุดข้อมูล
- หลังการเรียนรู้ ทำการทดสอบเพื่อวัดประสิทธิภาพของความรู้หรือรูปแบบที่เป็นผลจากการเรียนรู้
- ตัวอย่างเช่น naïve bayes, decision tree, neural networks
- ปัญหา การทำเฉลยของชุดฝึกข้อมูลใช้เวลานานมาก หรือขาดผู้เชี่ยวชาญ

การเรียนรู้แบบไม่มีผลเฉลย

- การเรียนรู้แบบไม่ต้องใช้ข้อมูลฝึกที่มีคำตอบ
- ไม่มีการบอกว่าผลลัพธ์เป็นอย่างไร
- เรียนรู้เพื่อหาความสัมพันธ์ของข้อมูล
- นิยมใช้ในการจัดกลุ่มข้อมูล แยกแยะข้อมูลเป็นกลุ่มๆ
- เป้าหมายการรวมกลุ่มข้อมูลที่มีคุณลักษณะใกล้เคียงกันไว้ในกลุ่มเดียวกัน
- ตัวอย่างเช่น Nearest neighbor classification, K-mean clustering, K-mode clustering, hierarchical clustering

การเรียนรู้แบบกึ่งมีผลเฉลย

- การเรียนรู้ที่สามารถเรียนรู้จากข้อมูลฝึกที่มีคำเฉลยบางตัวอย่าง
- การเรียนรู้ซับซ้อนกว่าการเรียนรู้แบบมีผลเฉลย
- ตัวอย่างเช่น Co-training algorithm

Naïve bayes classifier

ขั้นตอนวิธีการเรียนรู้

- ประยุกต์ใช้ทฤษฎีเบย์กับการจำแนกข้อมูล

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

- ถ้ามี attribute เดียวจะใช้สูตร

$$P(class | attribute) = \frac{P(class)P(attribute | class)}{P(attribute)}$$

- ถ้ามี attribute มากกว่า 1 ตัว จะใช้สูตร

$$P(class | a_1, a_2, \dots, a_n) = \frac{P(class)P(a_1, a_2, \dots, a_n | class)}{P(a_1, a_2, \dots, a_n)}$$

Naïve bayes classifier

ขั้นตอนวิธีการเรียนรู้

- ประยุกต์ใช้ทฤษฎีเบย์กับการจำแนกข้อมูล

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

- ถ้ามี attribute เดียวจะใช้สูตร

$$P(class | attribute) = \frac{P(class)P(attribute | class)}{P(attribute)}$$

- ถ้ามี attribute มากกว่า 1 ตัว จะใช้สูตร

$$P(class | a_1, a_2, \dots, a_n) = \frac{P(class)P(a_1, a_2, \dots, a_n | class)}{P(a_1, a_2, \dots, a_n)}$$

Naïve bayes classifier

- ค่าความน่าจะเป็นของเหตุการณ์ต่างๆ ร่วมกัน คำนวณได้โดย

$$P(a_1, a_2, \dots, a_n) = P(a_1)P(a_2 | a_1)..P(a_n | a_1, \dots, a_{n-1})$$

$$P(a_1, a_2, \dots, a_n | class) = P(a_1 | class)P(a_2 | a_1, class)..P(a_n | a_1, \dots, a_{n-1}, class)$$

- ค่าความน่าจะเป็นของ $P(a_1, a_2, \dots, a_n | class)$ หาได้ยาก ต้องใช้ข้อมูลฝึกจำนวนมากเพื่อให้ได้ค่าที่ถูกต้อง

- สมมติฐาน Naïve bayes พิจารณาในแง่ที่ attribute a_i เป็นอิสระต่อกัน ค่าความน่าจะเป็นการเกิดเหตุการณ์ต่างๆ ที่เป็นอิสระต่อกัน คำนวณได้โดย

$$P(a_1, a_2, \dots, a_n | class) = \prod_{i=1}^n P(a_i | class)$$

$$P(a_1, a_2, \dots, a_n | class) = P(a_1 | class)P(a_2 | class)..P(a_n | class)$$

ตัวอย่างการใช้ Naïve bayes classifier

X = (age = youth, income = medium, student = yes, credit = fair)

RID	age	income	student	credit	C_i : buy
1	youth	high	no	fair	C_2 : no
2	youth	high	no	excellent	C_2 : no
3	middle-aged	high	no	fair	C_1 : yes
4	senior	medium	no	fair	C_1 : yes
5	senior	low	yes	fair	C_1 : yes
6	senior	low	yes	excellent	C_2 : no
7	middle-aged	low	yes	excellent	C_1 : yes
8	youth	medium	no	fair	C_2 : no
9	youth	low	yes	fair	C_1 : yes
10	senior	medium	yes	fair	C_1 : yes
11	youth	medium	yes	excellent	C_1 : yes
12	middle-aged	medium	no	excellent	C_1 : yes
13	middle-aged	high	yes	fair	C_1 : yes
14	senior	medium	no	excellent	C_2 : no

ตัวอย่างการใช้ Naïve bayes classifier

$X = (\text{age} = \text{youth}, \text{income} = \text{medium}, \text{student} = \text{yes}, \text{credit} = \text{fair})$

- attributes ได้แก่ age, income, student, credit
- จำแนกเป็น 2 class (ลูกค้าจะซื้อคอมพิวเตอร์หรือไม่) ได้แก่ yes (class C_1) และ no (class C_2)
- ต้องการค่า $\text{classify}(\text{data}) = \arg \max_{c_j \in C} P(c_j) \prod_{i=1}^n P(a_i | c_j)$
- หาค่า $P(c_j)$ และ $P(a_i | c_j)$ เมื่อ $i = 1, 2$ และ $j=1,2,3,4$
- ได้แนวคิด/รูปแบบจากการเรียนรู้ ดังนี้

$$P(\text{buy} = \text{yes}) = \frac{9}{14}$$

$$P(\text{buy} = \text{no}) = \frac{5}{14}$$

$$P(\text{age} = \text{youth} | \text{buy} = \text{yes}) = \frac{2}{9}$$

$$P(\text{income} = \text{medium} | \text{buy} = \text{yes}) = \frac{4}{9}$$

$$P(\text{student} = \text{yes} | \text{buy} = \text{yes}) = \frac{6}{9}$$

$$P(\text{credit} = \text{fair} | \text{buy} = \text{yes}) = \frac{6}{9}$$

$$P(\text{age} = \text{youth} | \text{buy} = \text{no}) = \frac{3}{5}$$

$$P(\text{income} = \text{medium} | \text{buy} = \text{no}) = \frac{2}{5}$$

$$P(\text{student} = \text{yes} | \text{buy} = \text{no}) = \frac{1}{5}$$

$$P(\text{credit} = \text{fair} | \text{buy} = \text{no}) = \frac{2}{5}$$

ตัวอย่างการใช้ Naïve bayes classifier

$X = (\text{age} = \text{youth}, \text{income} = \text{medium}, \text{student} = \text{yes}, \text{credit} = \text{fair})$

$$\begin{aligned} P(X | \text{buy} = \text{yes}) &= P(\text{age} = \text{youth} | \text{buy} = \text{yes}) \\ &\quad P(\text{income} = \text{medium} | \text{buy} = \text{yes}) \\ &\quad P(\text{student} = \text{yes} | \text{buy} = \text{yes}) \\ &\quad P(\text{credit} = \text{fair} | \text{buy} = \text{yes}) \end{aligned} = \frac{2}{9} \frac{4}{9} \frac{6}{9} \frac{6}{9} = 0.044$$

$$P(X | \text{buy} = \text{no}) = \frac{3}{5} \frac{2}{5} \frac{1}{5} \frac{2}{5} = 0.019$$

- หาค่า max ระหว่างความน่าจะเป็นของ 2 class

$$P(\text{buy} = \text{yes})P(X | \text{buy} = \text{yes}) = 0.028$$

$$P(\text{buy} = \text{no})P(X | \text{buy} = \text{no}) = 0.007$$

- ดังนั้นจากข้อมูลที่ให้ X จึงทำนายว่าลูกค้ารายนี้มีจะซื้อคอมพิวเตอร์

ตัวอย่างการใช้ Naïve bayes จำแนกเอกสาร

	docID	words in document	in $c = \text{China?}$
training set	1	Chinese Beijing Chinese	yes
	2	Chinese Chinese Shanghai	yes
	3	Chinese Macao	yes
	4	Tokyo Japan Chinese	no
test set	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = 3/4 \text{ and } \hat{P}(\bar{c}) = 1/4$$

$$\hat{P}(\text{Chinese}|c) = (5 + 1)/(8 + 6) = 6/14 = 3/7$$

$$\hat{P}(\text{Tokyo}|c) = \hat{P}(\text{Japan}|c) = (0 + 1)/(8 + 6) = 1/14$$

$$\hat{P}(\text{Chinese}|\bar{c}) = (1 + 1)/(3 + 6) = 2/9$$

$$\hat{P}(\text{Tokyo}|\bar{c}) = \hat{P}(\text{Japan}|\bar{c}) = (1 + 1)/(3 + 6) = 2/9$$

$$\hat{P}(c|d_5) \propto 3/4 \cdot (3/7)^3 \cdot 1/14 \cdot 1/14 \approx 0.0003.$$

$$\hat{P}(\bar{c}|d_5) \propto 1/4 \cdot (2/9)^3 \cdot 2/9 \cdot 2/9 \approx 0.0001.$$

แบบฝึกหัด

ทำนายว่าวันนี้ นักกีฬาจะเล่นเทนนิสหรือไม่ เมื่อสถานการณ์ในวันนี้ เป็นดังนี้
Outlook=sunny, Temp=cool, Humid=high, Wind=strong

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

source: <http://jmvidal.cse.sc.edu/talks/bayesianlearning/nbex.xml>

Nearest neighbor classification

ขั้นตอนวิธีการเรียนรู้

- จัดกลุ่มโดยพิจารณาจากข้อมูลที่มีค่าใกล้เคียงกับค่าของข้อมูลที่พิจารณามากที่สุด
- ค่าใกล้เคียงอยู่ในรูปของ ระยะทาง/ระยะห่าง ระหว่างชุดข้อมูลกับข้อมูลที่สนใจ
- ระยะห่าง ระหว่างข้อมูลสามารถกำหนดได้ว่าต้องการเท่าใด ด้วยค่า k
- ถ้าจัดกลุ่มโดยใช้ระยะห่างเป็น 1 เรียกว่า 1-NN (One Nearest Neighbor)
- นิยมเรียกว่า k -NN โดย ค่า k คือระยะห่างที่กำหนด
- ใช้สมการ Euclidean ในการคำนวณหาระยะทาง

Nearest neighbor classification

- ใช้สมการ Euclidean ในการคำนวณหาระยะทาง

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

- p เป็นค่าในชุดข้อมูลที่ต้องการจัดกลุ่ม มีค่า $p_1, p_2, p_3, \dots, p_n$
- q เป็นค่าในชุดข้อมูลข้างเคียงที่นำมาพิจารณา มีค่า $q_1, q_2, q_3, \dots, q_n$

ตัวอย่างการใช้ k-NN

กำหนดสัตว์ 5 ชนิด จัดกลุ่มจากคุณสมบัติที่ใช้บ่งบอกสายพันธุ์

	เตี้ยเป็นหัว	ออกอุ้งเป็นไข่	เลี้ยงลูกด้วยนม	มีขนตามร่างกาย	ผิวหยาบ/มีเกล็ด	หูโตและหูแหลม	หูเล็กและหูกลม	ขนและขนนุ่มปกคลุม	หายใจด้วยเหงือก	สายพันธุ์
กบ		1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
เปิด		1	1	1		1		1		นก
ค่างคาว	1		1	1						สัตว์เลี้ยงลูกด้วยนม
งู		1			1		1			สัตว์เลี้ยงเลื้อยคลาน
ปลาแซลมอน		1			1		1		1	ปลา
สัตว์ปริศนา		1						1		???

ตัวอย่างการใช้ k-NN

ต้องการทราบว่าสัตว์ปริศนาใกล้เคียงกับสายพันธุ์ใดมากที่สุด (1NN)

- ระยะห่างของกบ (d_1)

$$d_1 = \sqrt{0 + (1-1)^2 + 0 + 0 + 0 + 0 + (0-1)^2 + (1-1)^2 + 0} = \sqrt{1}$$

- ระยะห่างของเป็ด (d_2)

$$d_2 = \sqrt{0 + (1-1)^2 + 0 + (0-1)^2 + 0 + (0-1)^2 + 0 + (1-1)^2 + 0} = \sqrt{2}$$

- ระยะห่างของค่างคาว (d_3)

$$d_3 = \sqrt{(0-1)^2 + (1-0)^2 + (0-1)^2 + (0-1)^2 + 0 + 0 + 0 + (1-0)^2} = \sqrt{5}$$

- ระยะห่างของงู (d_4)

$$d_4 = \sqrt{0 + (1-1)^2 + 0 + 0 + (0-1)^2 + 0 + (0-1)^2 + (1-0)^2 + 0} = \sqrt{3}$$

- ระยะห่างของปลาแซลมอน (d_5)

$$d_5 = \sqrt{0 + (1-1)^2 + 0 + 0 + (0-1)^2 + 0 + (0-1)^2 + (1-0)^2 + (0-1)^2} = \sqrt{4}$$

ตัวอย่างการใช้ k-NN

กำหนดสัตว์ 5 ชนิด จัดกลุ่มจากคุณสมบัติที่ใช้บ่งบอกสายพันธุ์

	ออกอุกเป็นหัว	ออกอุกเป็นไข่	เลี้ยงลูกด้วยนม	มีขนตามร่างกาย	ผิวหนัง/มีเกล็ด	สัตว์เลือดอุ่น	สัตว์เลือดเย็น	อยู่ในน้ำหรือใต้น้ำ	หายใจด้วยเหงือก	สายพันธุ์
กบ		1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
เปิด		1		1		1		1		นก
ค่างคาว	1		1	1						สัตว์เลี้ยงลูกด้วยนม
งู		1			1		1			สัตว์เลื้อยคลาน
ปลาแซลมอน		1			1		1		1	ปลา
จระเข้		1			1		1			สัตว์เลื้อยคลาน
แมว	1		1	1		1				สัตว์เลี้ยงลูกด้วยนม
นกกระจอกเทศ	1			1		1				นก
สัตว์ปริศนา			1	1		1		1		???

ตัวอย่างการใช้ k-NN

ต้องการทราบว่าสัตว์ปริศนาใกล้เคียงกับสายพันธุ์ใดมากที่สุด (3NN)

ระยะห่างของ

■ กบ = 2.449

■ ปลาแซลมอน = 2.645

■ เป็ด = 1.732

■ จระเข้ = 2.236

■ ค้างคาว = 1.414

■ แมว = 1

■ งู = 2.236

■ นกกระจอกเทศ = 1.414

จากระยะทางที่ได้จากข้อมูลทั้ง 8 ตัวอย่าง เลือกระยะทางที่น้อยสุด 3 อันดับ

1. แมว สัตว์เลี้ยงลูกด้วยนม
2. ค้างคาว สัตว์เลี้ยงลูกด้วยนม
3. นกกระจอกเทศ นก

แบบฝึกหัด

ต้องการทราบว่าสัตว์ปริศนาใกล้เคียงกับสายพันธุ์ใดมากที่สุด (1NN และ 3NN)

	เต่าเป็นหัว	เต่าเป็นใบ	เลี้ยงลูกด้วยนม	มีขนตามร่างกาย	ผิวหนัง/มีเกล็ด	สัตว์เลือดอุ่น	สัตว์เลือดเย็น	บินได้	หายใจด้วยเหงือก	สายพันธุ์
กบ		1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
ปลาหมอ		1			1		1		1	ปลา
จระเข้		1			1		1			สัตว์เลื้อยคลาน
ปลาหมอ					1		1		1	ปลา
เต่า		1					1			สัตว์เลื้อยคลาน
สุนัข	1		1	1		1				สัตว์เลื้อยคลาน
นกกระจอกเทศ	1			1		1				นก
สัตว์ปริศนา A	1			1						???
สัตว์ปริศนา B		1			1		1			???

บรรณานุกรม

- <http://cis.poly.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf>
- <http://jmvidal.cse.sc.edu/talks/bayesianlearning/nbex.xml>
- <http://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html>
- นवलวรรณ สุนทรภิชช์, 2552, ปัญญาประดิษฐ์, กรุงเทพฯ: ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.