COMP307 Week 7 (Tutorial)

- Announcements
  - Assignment 2
    Due: 23:59 Monday 8 May 2017
  - Assignment 3
    Due: 23:59 Monday 29 May
- Basic Rules
- Bayes Rules
- Naïve Bayes
  - Assumption
  - Why not directly calculate $P(clas|data)$?
  - Zero counting
- Conditionally independent VS fully independent

Rules

- Product Rule:
  P(X,Y)=P(X)*P(Y|X)
- Sum Rule:
  $$P(X)=\sum_y P(X, Y)$$
- Normalisation:
  $$\sum_x P(X)=1$$
  $$\sum_x P(X/Y)=1$$
- Independence
  - $\leftrightarrow$ P(X|Y) = P(X)
  - $\leftrightarrow$ P(X, Y) = P(X) * P(Y)

The Product Rule

Y \ X	A	B	C	
T	4	2	3	9
¬T	3	3	3	9
	7	5	6	***18***

- P(A)=7/18
- P(X=T) = 9/18
- P(X=T, Y=A) = 4/18
- P(X=T|Y=A) = 4/7
- P(Y=A|X=T) = 4/9

- P(X=T, Y=A) = P(X=T)*P(Y=A|X=T)

- ***The Product Rule: P(X,Y)=P(X)*P(Y|X)***

The Sum Rule

X \ Y	A	B	C	
T	4	2	3	9
¬T	3	3	3	9
	7	5	6	18

- P(X=T, Y=A) = 4/18
- P(X=T, Y=B) = 2/18
- P(X=T, Y=C) = 3/18
- P(X=T) = 9/18

- P(X=T) = P(X=T, Y=A) +P(X=T, Y=B) +P(X=T, Y=C)

- ***The Sum Rule***:

$$P(X) = \sum_y P(X, Y)$$

The Normalisation Rule

Y \ X	A	B	C	
T	4	2	3	9
¬T	3	3	3	9
	7	5	6	***18***

- P(X=T) = 9/18
- P(X=¬T) = 9/18
- P(Y=A|X=T) = 4/9
- P(Y=B|X=T) = 2/9
- P(Y=C|X=T) = 3/9

- P(X=T) +P(X=¬T) = 1
- P(Y=A|X=T) +P(Y=B|X=T) +P(Y=C|X=T) = 1
- ***The Normalisation Rule***:

$$\sum_x \mathrm{P(X)}{=}1$$

$$\sum_x \mathrm{P(X/Y)}{=}1$$

Example

- **W**indy or **C**alm
- **D**ay 1 ——>**D**ay 2
- P(D1=W) = 0.5
- P(D1=C) = 0.5
- P(D2=W|D1=W) = 0.6
- P(D2=C|D1=W) = 0.4
- P(D2=W|D1=C) = 0.3
- P(D2=C|D1=C) = 0.7

- Question: P(D2=W) ?
- Question: P(D3=W) ?
- Question: P(D3=C) ?

$$\boxed{\begin{aligned} P(D_2 = w) &= P(D_2 = w, D_1 = w) + P(D_2 = w, D_1 = c) \\ &= 0.3 + 0.15 \\ &= 0.45 \end{aligned}}$$

Bayes Rules

- P(A,B) = P(A|B) P(B)
- We can also get:
  P(A,B)= P(B|A) P(A)

Thomas Bayes (/ˈbeɪz/; c. 1701 – 7 April 1761)

- Bayes Rules:

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

- **More variables**

$$P(Y|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y)P(Y)}{P(X_1,\ldots,X_n)}$$

Bayes Rules for Classification

- Solution: First use Bayes' Law/Rules, calculate the probability of given instance belong to a class:

$$P(clas|data) = \frac{P(data|class) * P(class)}{P(data)}$$

- For example:

$$P(\mathrm{Reject} \,\middle|\, job = true \,\&\, dep = high \,\&\, fam = children)$$

$$= \frac{P(job = true \,\&\, dep = high \,\&\, fam = children|Reject) * P(\mathrm{Reject})}{P(job = true \,\&\, dep = high \,\&\, fam = children)}$$

P(***Reject***|job=true & dep = high & fam=children)

P(***Accept***|job=true & dep = high & fam=children)

Choose the highest probability

Naïve Bayes: Summary

1. Bayes Rules: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$ $\Rightarrow$ $P(Y|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y)P(Y)}{P(X_1,\ldots,X_n)}$

2. Classification: If *Y* is class label, *X1 .. Xn* features, the probability of an instance belong to a class is

$$P(Y|X_1,\ldots,X_n) = \frac{\boxed{P(X_1,\ldots,X_n|Y)}P(Y)}{P(X_1,\ldots,X_n)}$$

Too Hard

2. Assume features are conditionally independent: *given Y, X1 .. Xn are independent to each other:*

$$\boxed{P(X_1,\ldots,X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)}$$

$$P(clas|data) = \frac{P(data|class) * P(class)}{P(data)}$$

Naive Bayes

Chose the highest probability/Score

Bayes Rules for Classification

$$P(clas|data) = \frac{P(data|class) * \ P(class)}{P(data)}$$

- Why not directly calculate $P(clas|data)$?

P(***Reject***|job=true & dep = high & fam=children)

P(***Accept***|job=true & dep = high & fam=children)

Computing Probabilities: Counting Occurrences

	Approve	Reject
Class	5	5
Job=true	4	2
Job=false	1	3
dep=low	2	4
dep=high	3	1
fam=single	3	1
fam=couple	2	2
fam=children	0	2

	Approve	Reject
P(Class)	5/10	5/10
P(job=true|class)	4/5	2/5
P(job=false|class)	1/5	3/5
P(dep=low|class)	2/5	4/5
P(dep=high|class)	3/5	1/5
P(fam=single|class)	3/5	1/5
P(fam=couple|class)	2/5	2/5
P(fam=children|class)	0/5	2/5

Using Naive Bayes Classifier

$$P(\textbf{\textit{Reject}}|job=true \ \& \ dep = high \ \& \ fam=children)$$

$$= \frac{P(job=true \ \& \ dep = high \ \& \ fam=children|\textit{Reject}) \times P(\textit{Reject})}{P(job=true \ \& \ dep = high \ \& \ fam=children)}$$

$$= \frac{P\ (job=true|\textit{Reject}) \times P\ (dep=high|\textit{Reject}) \times P\ (fam=children|\textit{Reject}) \ \times P(Reject)}{P(job=true \ \& \ dep = high \ \& \ fam=children)}$$

$$= \frac{2/5\times1/5\times2/5\times1/2}{????}$$

$$P(\textbf{\textit{Accept}}|job=true \ \& \ dep = high \ \& \ fam=children)$$

$$= \frac{P(job=true \ \& \ dep = high \ \& \ fam=children|\textit{Accept}) \times P(\textit{Accept})}{P(job=true \ \& \ dep = high \ \& \ fam=children)}$$

$$= \frac{P\ (job=true|\textit{Accept}) \times P\ (dep=high|\textit{Accept}) \times P\ (fam=children|\textit{Accept}) \ \times P(\textit{Accept})}{P(job=true \ \& \ dep = high \ \& \ fam=children)}$$

$$= \frac{0}{????}$$

Dealing with Zero Counts

- Initialise table to contain small constant, e.g. 1
- This is not quite sound, but reasonable in practice

	Approve	Reject
Class	6	6
Job=true	5	3
Job=false	2	4
dep=low	3	5
dep=high	4	2
fam=single	4	2
fam=couple	3	3
fam=children	1	3

	Approve	Reject
P(Class)	6/12	6/12
P(job=true|class)	5/7	3/7
P(job=false|class)	2/7	4/7
P(dep=low|class)	3/7	5/7
P(dep=high|class)	4/7	2/7
P(fam=single|class)	4/8	2/8
P(fam=couple|class)	3/8	3/8
P(fam=children|class)	1/8	3/8

Compared with previous table, tricks here :
job and dep: 5+2=7; fam has 5+3=8;

Using Naive Bayes Classifier

P(***Reject***|job=true & dep = high & fam=children)

$$= \frac{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children|\mathit{Reject}) \times P(\mathit{Reject})}{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children)}$$

$$= \frac{P\ (job{=}true|\mathit{Reject}) \times P\ (dep{=}high|\mathit{Reject}) \times P\ (fam{=}children|\mathit{Reject}) \times P(Reject)}{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children)}$$

$$= \frac{3/7 \times 2/7 \times 3/8 \times 1/2}{????} = \frac{18/784}{????}$$

P(***Accept***|job=true & dep = high & fam=children)

$$= \frac{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children|\mathit{Accept}) \times P(\mathit{Accept})}{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children)}$$

$$= \frac{P\ (job{=}true|\mathit{Accept}) \times P\ (dep{=}high|\mathit{Accept}) \times P\ (fam{=}children|\mathit{Accept}) \times P(\mathit{Accept})}{P(job{=}true \ \& \ dep = high \ \& \ fam{=}children)}$$

$$= \frac{5/7 \times 4/7 \times 1/8 \times 1/2}{????} = \frac{20/784}{????}$$

Using Naive Bayes Classifier

Classify a new case: `(job=true & dep = high & fam=children)`

$$P(Reject|\text{job=true \& dep = high \& fam=children})$$

$$= \frac{P(\boxed{\text{job=true \& dep = high \& fam=children}}|Reject) \times P(Reject)}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{P(\boxed{\text{job=true}|Reject) \times P(\text{dep=high}|reject) \times P(\text{fam=children}|Reject}) \times P(Reject)}{P(\text{job=true}) \times P(\text{dep=high}) \times P(\text{fam=children})}$$

$$= \frac{0.4 \times 0.2 \times 0.4 \times 0.5}{\boxed{0.6 \times 0.3 \times 0.2}} = \frac{0.016}{0.036}$$

$$P(Accept|\text{job=true \& dep = high \& fam=children})$$

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|Accept) \times P(Accept)}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{0.8 \times 0.4 \times 0 \times 0.5}{\boxed{0.6 \times 0.3 \times 0.2}} = \frac{0}{0.036}$$

Using Naive Bayes Classifier

Classify a new case: `(job=true & dep = high & fam=children)`

$$P(Reject|\text{job=true \& dep = high \& fam=children})$$

$$= \frac{P(\boxed{\text{job=true \& dep = high \& fam=children}}|Reject) \times P(Reject)}{P(\text{job=true \& dep = high \& fam=children})}$$

A and B independent does not imply and is not implied by A and B are conditionally independent given C

$$P(Accept|\text{job=true \& dep = high \& fam=children})$$

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|Accept) \times P(Accept)}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{0.8 \times 0.4 \times 0 \times 0.5}{\boxed{0.6 \times 0.3 \times 0.2}} = \frac{0}{0.036}$$

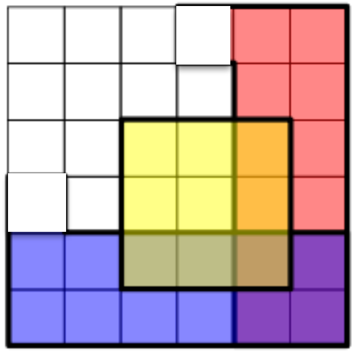
Conditional independence

- Two random variables X and Y are conditionally independent given a third random variable Z if and only if they are independent in their conditional probability distribution given Z.
- That is: X and Y are conditionally independent given Z if and only if, given any value of Z, the probability distribution of X is the same for all values of Y, and the probability distribution of Y is the same for all values of X.

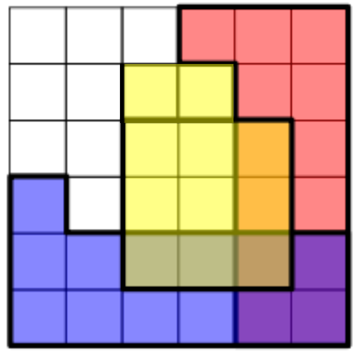
- **$X \perp Y$ neither implies nor is implied by $X \perp Y \mid Z$.**
  - **$P(X, Y|Z) = P(X|Z) * P(Y|Z)$**
  - **$P(X|Z) = P(X|Y,Z)$**

Independence VS Conditional Independence

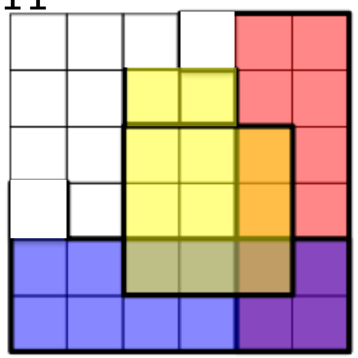
- R $\perp$ B, and R $\perp$ B | Y.
  - P(B)=P(R)=12/36=1/3, P(B,R)=4/36, equal P(B)*P(R)=1/9
  - P(B,R|Y)=1/9 equal P(B|Y)*P(R|Y)=3/9*3/9



- not R $\perp$ B, not R $\perp$ B| Y.
  - P(B)=P(R)=13/36, P(B,R)=4/36 not equal P(B)*P(R)
  - P(B,R|Y)=1/11, not equal P(B|Y)*P(R|Y)=3/11*3/11



- R $\perp$ B, not R $\perp$ B | Y.
  - P(B)=P(R)=12/36=1/3, P(B,R)=4/36, equal P(B)*P(R)=1/9
  - P(B,R|Y)=1/11, not equal P(B|Y)*P(R|Y)=3/11*3/11



- not R $\perp$ B, R $\perp$ B|Y.
  - P(B)=P(R)=13/36, P(B,R)=4/36 not equal P(B)*P(R)
  - P(B,R|Y)=1/9 equal P(B|Y)*P(R|Y)=3/9*3/9