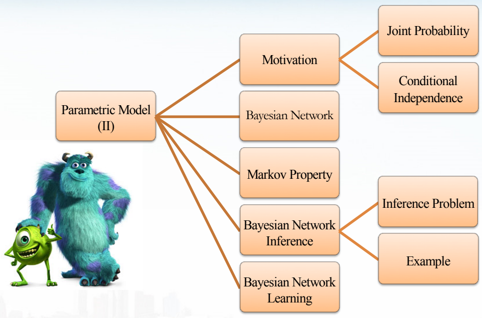
Lecture Notes on

Pattern Recognition

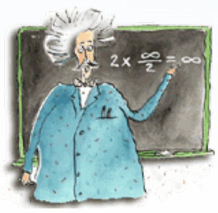
Session 04(B): Parametric Model (II)

Gang Li
School of Information Technology
Deakin University, VIC 3125, Australia

Road map



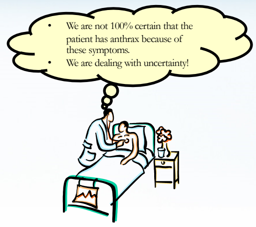
Motivation



- Jointed Probability
- Conditional Independence

Motivation Example

- Suppose you are trying to determine if a patient has inhalational anthrax.
- You observe the following symptoms:
  - The patient has a cough
  - The patient has a fever
  - The patient has difficulty breathing
- Now suppose you order an x-ray and observe that the patient has a wide mediastinum.
  - Your belief that that the patient is infected with inhalational anthrax is now much higher.



Reasoning with uncertainty

- In the previous slides, what you observed affected your belief that the patient is infected with anthrax
- Wouldn't it be nice if we had some methodology for reasoning with uncertainty?

The Joint Probability Distribution

- Joint probabilities can be between any number of variables:

$P(A = true, B = true, C = true)$

- For each combination of variables, we need to say how probable that combination is.
- The probabilities of these combinations need to sum to 1

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

The Joint Probability Distribution

- Once you have the joint probability distribution, you can calculate any probability involving *A*, *B*, and *C*
- For example:

P(A=true)
= sum of *P(A,B,C)* in rows with *A=true*

P(A=true, B = true | C=true)
= *P(A = true, B = true, C = true) / P(C = true)*

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

The Joint Probability Distribution

Limitations

- Lots of entries in the table to fill up!
- For *k* Boolean random variables, you need a table of size 2^k
- How do we use fewer numbers?
  – Need the concept of independence

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Independence

- Variables *A* and *B* are independent if any of the following hold:
  – *P(A,B) = P(A) P(B)*
  – *P(A| B) = P(A)*
  – *P(B| A) = P(B)*

This says that knowing the outcome of *A* does not tell me anything new about the outcome of *B*.

Independence

- How is independence useful?
  – Suppose you have n coin flips and you want to calculate the joint distribution $\boldsymbol{P}(C_1, \ldots, C_n)$
  – If the coin flips are not independent, you need **2ⁿ** values in the table
  – If the coin flips are independent, then

$$P(C_1,\ldots,C_n) = \prod_{i=1}^{n} P(C_i)$$

Each $P(C_i)$ table has 2 entries and there are n of them for a total of **2n** values

Conditional Independence

- Variables *A* and *B* are conditionally independent given *C* if any of the following hold:
  – *P(A, B | C) = P(A | C) P(B | C)*
  – *P(A | B, C) = P(A | C)*
  – *P(B | A, C) = P(B | C)*

Knowing C tells me everything about B. I don't gain anything by knowing A
- either because A doesn't influence B or
- because knowing C provides all the information knowing A would give

Bayesian Network

- What is BN?
- Example

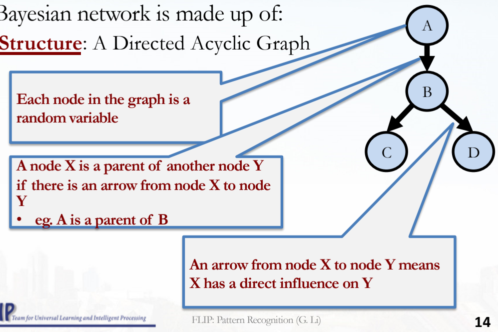
Reasoning with uncertainty

Judea Pearl

- 2011 winner of the ACM Turing Award, "*for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning*".

A Bayesian Network

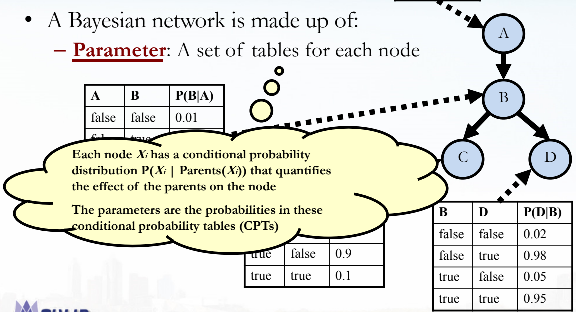
- A Bayesian network is made up of:
  - **Structure**: A Directed Acyclic Graph



A Bayesian Network

- A Bayesian network is made up of:
  - **Parameter**: A set of tables for each node

A	P(A)
false	0.6
true	0.4



B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

An Example

- You have a new burglar alarm installed at home.
  - It is fairly reliable at detecting burglary, but also sometimes responds to minor earthquakes.
- Your neighbors, Ali and Veli, promised to call you at work when they hear the alarm.
  - Ali always calls when he hears the alarm, but sometimes confuses telephone ringing with the alarm and calls too.
  - Veli likes loud music and sometimes misses the alarm.

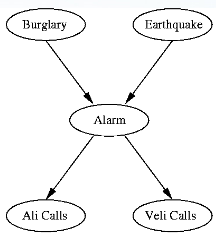
An Example

- Let us design the structure of the Bayesian net.
  - **Structure**: what are the main variables?
    - Alarm, Burglary, Earthquake, Ali calls, Veli calls
  - **Parameter**: what are the conditional dependencies among them?
    - Burglary (B) and earthquake (E) directly affect the probability of the alarm (A) going off
    - Whether or not Ali calls (AC) or Veli calls (VC) depends only on the alarm.

An Example

P(B=T)	P(B=F)
0.001	0.999

P(E=T)	P(E=F)
0.002	0.998



B	E	P(A=T)	P(A=F)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	P(AC=T)	P(AC=F)
T	0.90	0.10
F	0.05	0.95

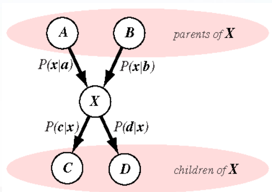
A	P(VC=T)	P(VC=F)
T	0.70	0.30
F	0.01	0.99

Markov Property

- Markov Property
- Joint Distribution Breakdown

Markov Property

- Each node is **conditionally independent** of its ancestors, given its parents



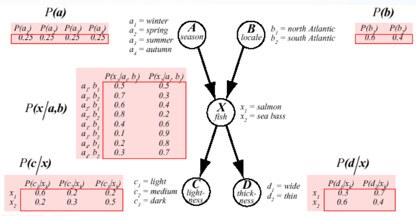
The Joint Probability Distribution

- Due to the Markov property, we can compute the joint probability distribution over all the variables $X_1, \ldots, X_n$ in the Bayesian net using the formula:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid Parents(X_i))$$

Another Bayesian Network (Joint Probability)

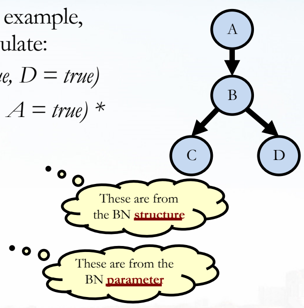
- $P(a_3, b_1, x_2, c_3, d_2) = P(a_3)P(b_1)P(x_2 / a_3, b_1)P(c_3 / x_2)P(d_2 / x_2)$
  $= 0.25 \times 0.6 \times 0.4 \times 0.5 \times 0.4 = 0.012$



A Bayesian Network (Joint Probability)

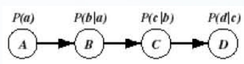
- Using the network in the example, suppose you want to calculate:

P(A = true, B = true, C = true, D = true)
*= P(A = true) * P(B = true | A = true) **
*P(C = true | B = true) **
P(D = true | B = true)
= (0.4)(0.3)*(0.1)*(0.95)*



Another Bayesian Network (The Probability at a Node)

- Determine the probability at D

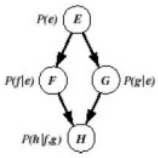


$$\begin{aligned} P(\mathrm{d}) &= \sum_{\mathrm{a,b,c}} P(\mathrm{a,b,c,d}) \\ &= \sum_{\mathrm{a,b,c}} P(\mathrm{a})P(\mathrm{b|a})P(\mathrm{c|b})P(\mathrm{d|c}) \\ &= \underbrace{\sum_{\mathrm{c}} P(\mathrm{d|c}) \underbrace{\sum_{\mathrm{b}} P(\mathrm{c|b}) \underbrace{\sum_{\mathrm{a}} P(\mathrm{b|a})P(\mathrm{a})}_{P(\mathrm{b})}}_{P(\mathrm{c})}}_{P(\mathrm{d})} \end{aligned}$$

Another Bayesian Network (The Probability at a Node)

- Determine the probability at H

$$
\begin{aligned}
P(\mathrm{h}) &= \sum_{\mathrm{e,f,g}} P(\mathrm{e,f,g,h}) \\
&= \sum_{\mathrm{e,f,g}} P(\mathrm{e})P(\mathrm{f}|\mathrm{e})P(\mathrm{g}|\mathrm{e})P(\mathrm{h}|\mathrm{f,g}) \\
&= \sum_{\mathrm{f,g}} P(\mathrm{h}|\mathrm{f,g}) \sum_{\mathrm{e}} P(\mathrm{e})P(\mathrm{f}|\mathrm{e})P(\mathrm{g}|\mathrm{e})
\end{aligned}
$$



Two basic BN problems

- **Evaluation (Inference)**
  - Given the model and the values of the observed variables, estimate the values of the hidden nodes.

    P(X | E)
- **Learning**
  - Given training data and prior information (e.g., expert knowledge, causal relationships),
    - **Structure Learning**: estimate the network structure, or
    - **Parameter Learning**: estimate the parameters of the probability

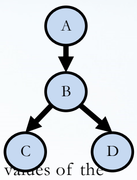
Bayesian Network Inference

- Inference Problem
- Example

Evaluation (Inference) Problem

- Let **X** denote the query variables and *e* denotes the evidence, then

$$P(\mathbf{X}/e) = \frac{P(\mathbf{X},e)}{P(e)} = \alpha P(\mathbf{X},e)$$



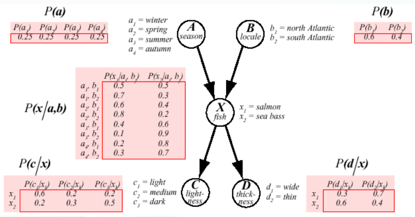
  - **Diagnosis/bottom-up reasoning**:
    - We observe the "leaves" and try to infer the values of the hidden causes
  - **Prediction/Top-Down reasoning**:
    - If we observe the "roots" and try to predict the effects

Evaluation (Inference) Problem

- **Exact inference** is an ***NP-hard*** problem because
  - the number of terms in the summations (integrals) for discrete (continuous) variables grows exponentially with increasing number of variables.
- **Approximate inference** methods have been popular:
  - sampling (Monte Carlo) methods
  - variational methods
  - loopy belief propagation
- Some restricted classes of networks, namely the singly connected networks where there is no more than one path between any two nodes, can be efficiently solved in time linear in the number of nodes.

A Bayesian Network (Inference)

- Classify a fish given that
  - the fish is light (c_1) and was caught in south Atlantic (b_2)
  - no evidence on the caught time nor its thickness.



A Bayesian Network (Inference)

$$\begin{aligned}
P(x_1|c_1,b_2) &= \frac{P(x_1,c_1,b_2)}{P(c_1,b_2)} \\
&= \alpha \sum_{\mathbf{a},\mathbf{d}} P(x_1,\mathbf{a},b_2,c_1,\mathbf{d}) \\
&= \alpha \sum_{\mathbf{a},\mathbf{d}} P(\mathbf{a})P(b_2)P(x_1|\mathbf{a},b_2)P(c_1|x_1)P(\mathbf{d}|x_1) \\
&= \alpha P(b_2)P(c_1|x_1) \\
&\quad \times \left[\sum_{\mathbf{a}} P(\mathbf{a})P(x_1|\mathbf{a},b_2)\right]\left[\sum_{\mathbf{d}} P(\mathbf{d}|x_1)\right]
\end{aligned}$$

$$\begin{aligned}
&= \alpha P(b_2)P(c_1|x_1) \\
&\quad \times [P(a_1)P(x_1|a_1,b_2) + P(a_2)P(x_1|a_2,b_2) + P(a_3)P(x_1|a_3,b_2) + P(a_4)P(x_1|a_4,b_2)] \\
&\quad \times \underbrace{[P(d_1|x_1) + P(d_2|x_1)]}_{=1} \\
&= \alpha(0.4)(0.6)\,[(0.25)(0.7) + (0.25)(0.8) + (0.25)(0.1) + (0.25)(0.3)]\,1.0 \\
&= \alpha\ 0.114.
\end{aligned}$$

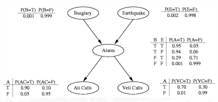
A Bayesian Network (Inference)

- Similarly $P(x_2/c_1,b_2)=a\ 0.066$
- Normalize probabilities (not needed necessarily):
  - $P(x_1/c_1,b_2)+P(x_2/c_1,b_2)=1 \quad (a=1/0.18)$
  - $P(x_1/c_1,b_2)= 0.63$
  - $P(x_2/c_1,b_2)= 0.27$

Another Bayesian Network (Inference)

- What is the probability that the alarm sounded but neither a burglary nor an earthquake, and both Ali and Veli call?

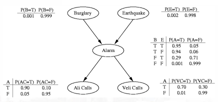
$$\begin{aligned}
&P(AC,VC,A,\neg B,\neg E) \\
&= P(AC|A)P(VC|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E) \\
&= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \\
&= 0.00062
\end{aligned}$$



Another Bayesian Network (Inference)

- What is the probability that there is a burglary given that Ali calls?

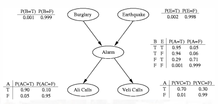
$$\begin{aligned}
P(B|AC) &= \frac{P(B,AC)}{P(AC)} \\
&= \frac{\sum_{vc}\sum_{a}\sum_{e} P(AC|a)P(vc|a)P(a|B,e)P(B)P(e)}{P(B,AC)+P(\neg B,AC)} \\
&= \frac{0.00084632}{0.00084632+0.0513} \\
&= 0.0162
\end{aligned}$$



Another Bayesian Network (Inference)

- What is the probability that there is a burglary given that Veli also calls right after Ali hangs up?

$$P(B|AC,VC) = \frac{P(B,AC,VC)}{P(AC,VC)} = 0.29$$



Bayesian Network Learning

- Learning B.N.
- N.B.C

Learning Bayesian Network

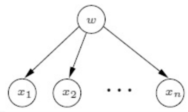
- The simplest situation is the one where the network structure is completely known
  - either specified by an expert or
  - designed using causal relationships between the variables
- Other situations with increasing complexity are:
  - known structure but unobserved variables
  - unknown structure with observed variables, and
  - unknown structure with unobserved variables.

Structure	Observability: Full	Observability: Partial
Known	Maximum Likelihood Estimation	EM (or gradient ascent)
Unknown	Search through model space	EM + search through model space

Naïve Bayes Net

- When dependency relationships among features are unknown, we can assume that features are conditionally independent given the class:

$$p(x_1, \ldots, x_n | w) = \prod_{i=1}^{n} p(x_i | w)$$



Naïve Bayes Classifier (NBC)

- NBC Training
- Prediction with NBC

Play Football: Training Data Set

- Consider a decision to *Play* outdoor or not [Weka, ch4]
  - Suppose we've collected the data for the past 2 weeks

Day	Outlook	Temperature	Humidity	Wind	Play Football
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	Weak	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Play Football: Training Data Set

- For a new day with following measures for Outlook, Temperature, Humidity and Windy, what is the prediction for Play?

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

  - We can answer question like this with many different methods
    - Naïve Bayes Classifier (NBC)
    - or Decision Trees

Meet the NBC

- It is a **Bayesian** method
  - A simple but very effective classification method
  - Training data can be discarded once trained.
  - Follows a typical setting for a supervised learning method in machine learning
    - Step 1: train the model using training data (e.g., MLE)
    - Step 2: use trained model to make prediction

Train an NBC

This is the feature X_1 | This is the feature X_2 | This is the feature X_3 | This is the feature X_4 | This is the class label Y

Day	Outlook	Temperature	Humidity	Wind	Play Football
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	Weak	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Train an NBC

- We need introduce three concepts:
  - Marginal probability
  - Joint probability
  - and Conditional probability

Train an NBC

- Marginal probability
  - What is the probability Pr(Outlook = sunny)?

Outlook
Sunny
Sunny
Overcast
Rain
Rain
Rain
Overcast
Sunny
Sunny
Rain
Sunny
Overcast
Overcast
Rain

Method 1: use the frequency table

Outlook	Frequency
sunny	5
overcast	4
rain	5

Pr(Outlook=sunny) $= \frac{5}{14}$

Pr(Outlook=overcast) $= \frac{4}{14}$

Pr(Outlook=rain) $= \frac{5}{14}$

Outlook	Probability
sunny	5/14
overcast	4/14
rain	5/14

Marginal probability for Outlook

Train an NBC

- Marginal probability
  - What is the probability Pr(Play = yes)?

Play Football
No
No
Yes
Yes
Yes
No
Yes
No
Yes
Yes
Yes
Yes
Yes
No

Method 1: use the frequency table

Play	Frequency
yes	9
no	5

Pr(Play=yes) $= \frac{9}{14}$

Pr(Play=no) $= \frac{5}{14}$

Play	Probability
yes	9/14
no	5/14

Marginal probability for Play

Train an NBC

- Joint probability Pr(Outlook = sunny, Play = yes)
  - Construct the contingency table for Outlook and Play

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Outlook \ Play	yes	no
sunny	2	3
overcast	4	0
rainy	3	2

Pr(Outlook = sunny, Play = yes) $= \frac{2}{14}$

Pr(Outlook = sunny, Play = no) $= \frac{3}{14}$

Pr(Outlook = overcast, Play = yes) $= \frac{4}{14}$

….

Train an NBC

- Joint probability Pr(Outlook = sunny, Play = yes)
  - Construct the contingency table for Outlook and Play

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Outlook \ Play	yes	no
sunny	2	3
overcast	4	0
rainy	3	2

Contingency table

Outlook \ Play	yes	no
sunny	2/14	3/14
overcast	4/14	0
rainy	3/14	2/14

Joint probability distribution for Outlook and Play

Train an NBC

DEAKIN Worldly

- Marginal probability Pr(Outlook = sunny)
  - the marginalization rule: $\Pr(X) = \sum_y P(X, Y = y)$

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

$\Pr(\text{Outlook=rainy}) = \frac{3}{14} + \frac{2}{14} = \frac{5}{14}$

Play

Outlook	yes	no
sunny	2/14	3/14
overcast	4/14	0
rainy	3/14	2/14

$\Pr(\text{Outlook=sunny})$

$= \Pr(\text{Outlook=sunny, Play=yes}) + \Pr(\text{Outlook=sunny, Play=no})$

$= \frac{2}{14} + \frac{3}{14} = \frac{5}{14}$

Outlook	Probability
sunny	5/14
overcast	4/14
rain	5/14

Marginal probability for Outlook computed before

Train an NBC

DEAKIN Worldly

- Conditional probability Pr(Outlook=sunny | Play = yes)
  - Construct the contingency table for Outlook and Play

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Play

Outlook	yes	no
sunny	2	3
overcast	4	0
rainy	3	2

$\Pr(\text{Outlook} = \text{sunny} \mid \text{Play} = \text{yes}) = \frac{2}{9}$

$\Pr(\text{Outlook} = \text{overcast} \mid \text{Play} = \text{yes}) = \frac{4}{9}$

$\Pr(\text{Outlook} = \text{rainy} \mid \text{Play} = \text{yes}) = \frac{3}{9}$

Train an NBC

DEAKIN Worldly

- Conditional probability Pr(Outlook=sunny | Play = yes)
  - Construct the contingency table for Outlook and Play

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Play

Outlook	yes	no
sunny	2	3
overcast	4	0
rainy	3	2

$\Pr(\text{Outlook} = \text{sunny} \mid \text{Play} = \text{no}) = \frac{3}{5}$

$\Pr(\text{Outlook} = \text{overcast} \mid \text{Play} = \text{no}) = 0$

$\Pr(\text{Outlook} = \text{rainy} \mid \text{Play} = \text{no}) = \frac{2}{5}$

Train an NBC

DEAKIN Worldly

- Conditional probability Pr(Outlook=sunny | Play = yes)
  - Construct the contingency table for Outlook and Play

Outlook	Play Football
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Play

Outlook	yes	no
sunny	2	3
overcast	4	0
rainy	3	2

Play

Outlook	yes	no
sunny	2/9	3/5
overcast	4/9	0
rainy	3/9	2/5

conditional

$\Pr(\text{Outlook} = \text{sunny} \mid \text{Play} = \text{no}) = \frac{3}{5}$

$\Pr(\text{Outlook} = \text{overcast} \mid \text{Play} = \text{no}) = 0$

$\Pr(\text{Outlook} = \text{rainy} \mid \text{Play} = \text{no}) = \frac{2}{5}$

Train an NBC

DEAKIN Worldly

- Training = conditional distribution for class Play

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	Str.	2/9	3/5	9/14	5/14
overcast	4/9	0	mild	4/9	2/5	normal	6/9	1/5	weak	7/9	2/5		
rainy	3/9	2/5	cold	3/9	1/5								

Marginal or prior for Play

Prediction with NBC

DEAKIN Worldly

- Predict Play outcome for a new day

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

- This amounts to compute the conditional probability:
  - $\Pr(P = yes | O = overcast, T = hot, H = normal, W = weak)$

versus

  - $\Pr(P = no | O = overcast, T = hot, H = normal, W = weak)$
- If the probability for P=yes > P=no, then predict yes, else predict no.

Prediction with NBC

- Predict Play outcome for a new day

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

- How to calculate the class conditional probability?
  - $\Pr(P = yes | O = overcast, T = hot, H = normal, W = weak)$
- We need the Bayes rule

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

constant w.r.t. θ

$$\underbrace{p(\theta \mid D)}_{posterior} \propto \underbrace{p(\theta)}_{prior} \times \underbrace{p(D \mid \theta)}_{likelihood}$$

Prediction with NBC

- Predict Play outcome for a new day

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

- How to calculate the class conditional probability?

$\Pr(P = yes | O = overcast, T = hot, H = normal, W = weak)$

$\propto \Pr(O = overcast, T = hot, H = normal, W = no \mid P = yes) \Pr(P = yes)$

$$\propto \underbrace{\Pr(O = overcast \mid P = yes) \Pr(T = hot \mid P = yes) \Pr(H = normal \mid P = yes) \Pr(W = no \mid P = yes)}_{\text{Independently factorized = Naïve assumption}} \Pr(P = yes)$$

Bayes'rule + Naïve assumption = Naïve Bayes Model

Prediction with NBC

- Predict Play outcome for a new day

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

- How to calculate the class conditional probability?

$\Pr(P = yes | O = overcast, T = hot, H = normal, W = weak)$

$\propto \Pr(O = overcast, T = hot, H = normal, W = no \mid P = yes) \Pr(P = yes)$

$\propto \Pr(O = overcast \mid P = yes) \Pr(T = hot \mid P = yes) \Pr(H = normal \mid P = yes) \Pr(W = no \mid P = yes) \Pr(P = yes)$

$\propto \frac{4}{9}\frac{2}{9}\frac{6}{9}\frac{7}{9}\frac{9}{14}$

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	$\boxed{2/9}$	2/5	high	3/9	4/5	Str.	2/9	3/5	$\boxed{9/14}$	5/14
overcast	$\boxed{4/9}$	0	mild	4/9	2/5	normal	$\boxed{6/9}$	1/5	weak	$\boxed{7/9}$	2/5		
rainy	3/9	2/5	cold	3/9	1/5								

Prediction with NBC

- Predict Play outcome for a new day

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

- How to calculate the class conditional probability?

$\Pr(P = yes | O = overcast, T = hot, H = normal, W = weak)$

$\Pr(P = no | O = overcast, T = hot, H = normal, W = no) \propto 0$

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	2/9	$\boxed{2/5}$	high	3/9	4/5	Str.	2/9	3/5	9/14	$\boxed{5/14}$
overcast	4/9	$\boxed{0}$	mild	4/9	2/5	normal	6/9	$\boxed{1/5}$	weak	7/9	$\boxed{2/5}$		
rainy	3/9	2/5	cold	3/9	1/5								

Prediction with NBC

- When data is insufficient, NBC may 'overfit' what it sees
- This probability will always zero:

$\Pr(P = no | O = overcast, T = ?, H = ?, W = ?) = 0$

  - Why? because $\Pr(H = no | O = overcast) = 0$
- How to resolve this?
  - Answer: add a pseudo-count, e.g., 1 to every entry.

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	Str.	2/9	3/5	9/14	5/14
overcast	4/9	$\boxed{0}$	mild	4/9	2/5	normal	6/9	1/5	weak	7/9	2/5		
rainy	3/9	2/5	cold	3/9	1/5								

Prediction with NBC

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	3/12	4/8	hot	3/12	3/8	high	4/11	5/7	Str.	3/11	4/7	10/16	6/16
overcast	5/12	1/8	mild	5/12	3/8	normal	7/11	2/7	weak	8/11	3/7		
rainy	4/12	3/8	cold	4/12	2/8								

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	Str.	2/9	3/5	9/14	5/14
overcast	4/9	$\boxed{0}$	mild	4/9	2/5	normal	6/9	1/5	weak	7/9	2/5		
rainy	3/9	2/5	cold	3/9	1/5								

Prediction with NBC

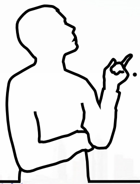
Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	3/12	4/8	hot	3/12	3/8	high	4/11	5/7	Str.	3/11	4/7	10/16	6/16
overcast	5/12	1/8	mild	5/12	3/8	normal	7/11	2/7	weak	8/11	3/7		
rainy	4/12	3/8	cold	4/12	2/8								

After adding pseudo-count to avoid overfitting, what is the prediction for Play now?

Day	Outlook	Temperature	Humidity	Wind	Play Football
D15	Overcast	Hot	Normal	Weak	????

TULIP Team for Universal Learning and Intelligent Processing   FLIP: Pattern Recognition (G. Li)   61

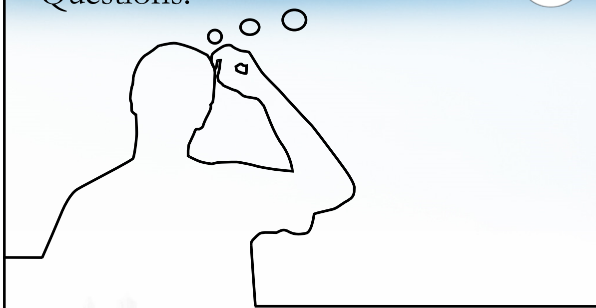
Seminar S06



- Topic
  - Try to model a problem in your daily life using **Bayesian Network**. Based on it, solve **AT LEAST 3 QUESTIONS**
    - including both joint probability, and inference.
- Requirements
  - Prepare a **15 minutes** talk on your chosen topic
  - Make **ppt** to assist your talk
  - Prepare **at least 3 questions** to ask the audience after your talk
  - Get ready to **take questions** from the audience
- Hints
  - Try **challenging** and **interesting** problems
  - Do statistics to get the probabilities in the B.N.
  - **You can make reasonable assumptions**

TULIP Team for Universal Learning and Intelligent Processing   FLIP: Pattern Recognition (G. Li)   62

Questions?



TULIP Team for Universal Learning and Intelligent Processing   FLIP: Pattern Recognition (G. Li)   63