

Note to other teachers and users of these slides.  
Andrew would be delighted if you found this source material useful in giving your own lectures. Feel free to use these slides verbatim, or to modify them to fit your own needs. PowerPoint originals are available. If you make use of a significant portion of these slides in your own lecture, please include this message, or the following link to the source repository of Andrew's tutorials: <http://www.cs.cmu.edu/~awm/tutorials>. Comments and corrections gratefully received.

# Information Gain

**Andrew W. Moore**  
**Professor**  
**School of Computer Science**  
**Carnegie Mellon University**

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

[awm@cs.cmu.edu](mailto:awm@cs.cmu.edu)

412-268-7599

Copyright © 2001, 2003, Andrew W. Moore

## Bits

You are watching a set of independent random samples of  $X$

You see that  $X$  has four possible values

|                |                |                |                |
|----------------|----------------|----------------|----------------|
| $P(X=A) = 1/4$ | $P(X=B) = 1/4$ | $P(X=C) = 1/4$ | $P(X=D) = 1/4$ |
|----------------|----------------|----------------|----------------|

So you might see: BAACBADCDADDDA...

You transmit data over a binary serial link. You can encode each reading with two bits (e.g.  $A = 00$ ,  $B = 01$ ,  $C = 10$ ,  $D = 11$ )

0100001001001110110011111100...

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 2

ملاحظة للمدرسين والمستخدمين الآخرين لهذه الشرائح. سيسعد أندرو إذا وجدت هذا المصدر مفيداً في إلقاء محاضراتك الخاصة. لا تتردد في استخدام هذه الشرائح حرفياً ، أو تعديلها لتناسب احتياجاتك الخاصة. أصول PowerPoint متوفرة. إذا كنت تستخدم جزءاً كبيراً من هذه الشرائح في محاضرتك الخاصة ، فيرجى تضمين هذه الرسالة ، أو الرابط التالي إلى مستودع المصدر لدروس أندرو التعليمية: <http://www.cs.cmu.edu/~awm/tutorials> . التعليقات والتصحيحات الواردة بامتنان.

# كسب المعلومات

أندرو ديليو مور

أستاذ

كلية علوم الكمبيوتر بجامعة

كارنيجي ميلون

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

[awm@cs.cmu.edu](mailto:awm@cs.cmu.edu)

412-268-7599

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## ب

أنت تشاهد مجموعة من العينات العشوائية المستقلة لـ  $X$

تري أن  $X$  لديها أربع قيم محتملة

|                  |                  |                  |                  |
|------------------|------------------|------------------|------------------|
| $P(X = D) = 1/4$ | $P(X = C) = 1/4$ | $P(X = B) = 1/4$ | $P(X = A) = 1/4$ |
|------------------|------------------|------------------|------------------|

لذلك قد ترى: BAACBADCDADDDA ...

تقوم بنقل البيانات عبر ارتباط تسلسلي ثنائي. يمكنك تشفير كل قراءة بتين (على سبيل المثال ،  $D = 11$  ،  $C = 10$  ،  $B = 01$  ،  $A = 00$ )

... 0100001001001110110011111100

كسب المعلومات: الشريحة 2

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## Fewer Bits

Someone tells you that the probabilities are not equal

|                |                |                |                |
|----------------|----------------|----------------|----------------|
| $P(X=A) = 1/2$ | $P(X=B) = 1/4$ | $P(X=C) = 1/8$ | $P(X=D) = 1/8$ |
|----------------|----------------|----------------|----------------|

It's possible...

...to invent a coding for your transmission that only uses 1.75 bits on average per symbol. How?

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 3

## Fewer Bits

Someone tells you that the probabilities are not equal

|                |                |                |                |
|----------------|----------------|----------------|----------------|
| $P(X=A) = 1/2$ | $P(X=B) = 1/4$ | $P(X=C) = 1/8$ | $P(X=D) = 1/8$ |
|----------------|----------------|----------------|----------------|

It's possible...

...to invent a coding for your transmission that only uses 1.75 bits on average per symbol. How?

|   |     |
|---|-----|
| A | 0   |
| B | 10  |
| C | 110 |
| D | 111 |

(This is just one of several ways)

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 4

## بت أقل

يخبرك أحدهم أن الاحتمالات ليست متساوية

|                  |                  |                  |                  |
|------------------|------------------|------------------|------------------|
| $P(X = D) = 1/8$ | $P(X = C) = 1/8$ | $P(X = B) = 1/4$ | $P(X = A) = 1/2$ |
|------------------|------------------|------------------|------------------|

انه ممكن...

...لاختراع تشفير للإرسال الخاص بك يستخدم فقط 1.75 بت في المتوسط لكل رمز. كيف؟

كسب المعلومات: الشريحة 3

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## بت أقل

يخبرك أحدهم أن الاحتمالات ليست متساوية

|                  |                  |                  |                  |
|------------------|------------------|------------------|------------------|
| $P(X = D) = 1/8$ | $P(X = C) = 1/8$ | $P(X = B) = 1/4$ | $P(X = A) = 1/2$ |
|------------------|------------------|------------------|------------------|

انه ممكن...

...لاختراع تشفير للإرسال الخاص بك يستخدم فقط 1.75 بت في المتوسط لكل رمز. كيف؟

|     |   |
|-----|---|
| 0   | أ |
| 10  | ب |
| 110 | ج |
| 111 | د |

(هذه مجرد طريقة من عدة طرق)

كسب المعلومات: الشريحة 4

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## Fewer Bits

Suppose there are three equally likely values...

|                |                |                |
|----------------|----------------|----------------|
| $P(X=A) = 1/3$ | $P(X=B) = 1/3$ | $P(X=C) = 1/3$ |
|----------------|----------------|----------------|

Here's a naïve coding, costing 2 bits per symbol

|   |    |
|---|----|
| A | 00 |
| B | 01 |
| C | 10 |

Can you think of a coding that would need only 1.6 bits per symbol on average?

In theory, it can in fact be done with 1.58496 bits per symbol.

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 5

## General Case

Suppose  $X$  can have one of  $m$  values...  $V_1, V_2, \dots, V_m$

|                  |                  |      |                  |
|------------------|------------------|------|------------------|
| $P(X=V_1) = p_1$ | $P(X=V_2) = p_2$ | .... | $P(X=V_m) = p_m$ |
|------------------|------------------|------|------------------|

What's the smallest possible number of bits, on average, per symbol, needed to transmit a stream of symbols drawn from  $X$ 's distribution? It's

$$H(X) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_m \log_2 p_m$$

$$= -\sum_{j=1}^m p_j \log_2 p_j$$

$H(X)$  = The entropy of  $X$

- "High Entropy" means  $X$  is from a uniform (boring) distribution
- "Low Entropy" means  $X$  is from varied (peaks and valleys) distribution

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 6

## بت أقل

افتراض أن هناك ثلاث قيم متساوية الاحتمال ...

|                          |                          |                          |
|--------------------------|--------------------------|--------------------------|
| الفوسفور (X = A) $1/3 =$ | الفوسفور (X = B) $1/3 =$ | الفوسفور (X = C) $1/3 =$ |
|--------------------------|--------------------------|--------------------------|

هذا ترميز ساذج ، يكلف 2 بت لكل رمز

|   |    |
|---|----|
| أ | 00 |
| ب | 01 |
| ج | 10 |

هل يمكنك التفكير في ترميز يحتاج فقط 1.6 بت لكل رمز في المتوسط؟

من الناحية النظرية ، يمكن في الواقع أن يتم ذلك باستخدام 1.58496 بت لكل رمز.

كسب المعلومات: الشريحة 5

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## الحالة العامة

افتراض أن X يمكن أن يكون له واحد من القيم...الخامس<sub>1</sub>، الخامس<sub>2</sub>،...الخامس<sub>M</sub>

|                |                |      |                |
|----------------|----------------|------|----------------|
| $V_1 = (P_1)X$ | $V_2 = (P_2)X$ | .... | $V_M = (P_M)X$ |
|----------------|----------------|------|----------------|

ما هو أصغر عدد ممكن من البتات ، في المتوسط ، لكل الرمز اللازم لنقل دفق من الرموز المستمدة من توزيع X؟ إنه

$$H(X) = - \sum_{i=1}^M P_i \log_2 P_i$$

$$= - \sum_{i=1}^M P_i \log_2 P_i$$

X إنتروبيا H(X)

- "إنتروبيا عالية" تعني أن X ناتجة عن توزيع موحد (ممل)
- يعني "إنتروبيا منخفض" أن X من توزيع (قمم ووديان) متنوع

كسب المعلومات: الشريحة 6

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## General Case

Suppose  $X$  can have one of  $m$  values...  $V_1, V_2, \dots, V_m$

|                  |                  |      |                  |
|------------------|------------------|------|------------------|
| $P(X=V_1) = p_1$ | $P(X=V_2) = p_2$ | .... | $P(X=V_m) = p_m$ |
|------------------|------------------|------|------------------|

What's the smallest possible number of symbols,  $H(X)$ , to represent  $X$ 's distribution?

A histogram of the frequency distribution of values of  $X$  would be flat

A histogram of the frequency distribution of values of  $X$  would have many lows and one or two highs

$$H(X) = -\sum_{j=1}^m p_j \log_2 p_j$$

$H(X)$  = The entropy of  $X$

- "High Entropy" means  $X$  is from a uniform (boring) distribution
- "Low Entropy" means  $X$  is from varied (peaks and valleys) distribution

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 7

## General Case

Suppose  $X$  can have one of  $m$  values...  $V_1, V_2, \dots, V_m$

|                  |                  |      |                  |
|------------------|------------------|------|------------------|
| $P(X=V_1) = p_1$ | $P(X=V_2) = p_2$ | .... | $P(X=V_m) = p_m$ |
|------------------|------------------|------|------------------|

What's the smallest possible number of symbols,  $H(X)$ , to represent  $X$ 's distribution?

A histogram of the frequency distribution of values of  $X$  would be flat

A histogram of the frequency distribution of values of  $X$  would have many lows and one or two highs

..and so the values sampled from it would be all over the place

..and so the values sampled from it would be more predictable

$H(X)$  = The entropy of  $X$

- "High Entropy" means  $X$  is from a uniform (boring) distribution
- "Low Entropy" means  $X$  is from varied (peaks and valleys) distribution

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 8





## Entropy in a nut-shell



Low Entropy



High Entropy

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 9

## Entropy in a nut-shell



Low Entropy

..the values (locations of soup) sampled entirely from within the soup bowl



High Entropy

..the values (locations of soup) unpredictable... almost uniformly sampled throughout our dining room

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 10

## الانتروبيا في قشرة الجوز



ارتفاع الانتروبيا



إنتروبيا منخفضة

كسب المعلومات: الشريحة 9

حقوق النشر © 2001 ، أندرو ديليو مور

## الانتروبيا في قشرة الجوز



ارتفاع الانتروبيا



إنتروبيا منخفضة

. القيم (مواقع الحساء) غير متوقعة ...  
يتم أخذ عينات بشكل موحد تقريباً  
في جميع أنحاء غرفة الطعام لدينا

. القيم (مواقع الحساء) التي  
تم أخذ عينات منها  
كليا من الداخل  
وعاء الحساء

كسب المعلومات: الشريحة 10

حقوق النشر © 2001 ، أندرو ديليو مور

## Specific Conditional Entropy $H(Y|X=v)$

Suppose I'm trying to predict output  $Y$  and I have input  $X$

$X$  = College Major

$Y$  = Likes "Gladiator"

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

Let's assume this reflects the true probabilities

E.G. From this data we estimate

- $P(\text{LikeG} = \text{Yes}) = 0.5$
- $P(\text{Major} = \text{Math} \ \& \ \text{LikeG} = \text{No}) = 0.25$
- $P(\text{Major} = \text{Math}) = 0.5$
- $P(\text{LikeG} = \text{Yes} \mid \text{Major} = \text{History}) = 0$

Note:

- $H(X) = 1.5$
- $H(Y) = 1$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 11

## Specific Conditional Entropy $H(Y|X=v)$

$X$  = College Major

$Y$  = Likes "Gladiator"

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

**Definition of Specific Conditional Entropy:**

$H(Y|X=v)$  = The entropy of  $Y$  among only those records in which  $X$  has value  $v$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 12

## الانتروبيا الشرطية المحددة $H(Y|X) = v$

لنفترض أنني أحاول التنبؤ بالمخرج  $Y$  ولدي إدخال  $X$

لنفترض أن هذا يعكس الاحتمالات الحقيقية

تخصص الكلية  $X =$

"Gladiator" لا يكت  $Y =$

من هذه البيانات نقدر EG

•  $0.5 = \text{نعم} = P(\text{Like} | G)$

•  $0.25 = \text{لا} = P(\text{Like} | G) \& \text{الرأي} = \text{الرياضيات} = P$

•  $0.5 = \text{الرأي} = \text{الرياضيات} = P$

•  $0 = \text{نعم} = P(\text{Like} | G) \text{التخصص} = \text{التاريخ} = 0$

ملحوظة:

•  $1.5 = \text{ح(س)}$

•  $1 = \text{ح(ص)}$

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 11

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## الانتروبيا الشرطية المحددة $H(Y|X) = v$

تعريف الانتروبيا الشرطية المحددة:

"Gladiator" إعجاب  $Y =$

$X = \text{College Major}$

ح(ص | س = ت) = إنتروبيا ص من بين تلك السجلات فقط التي  $X$  له قيمة الخامسة

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 12

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## Specific Conditional Entropy $H(Y|X=v)$

X = College Major  
Y = Likes "Gladiator"

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

**Definition of Specific Conditional Entropy:**

$H(Y|X=v)$  = The entropy of  $Y$  among only those records in which  $X$  has value  $v$

**Example:**

- $H(Y|X=Math) = 1$
- $H(Y|X=History) = 0$
- $H(Y|X=CS) = 0$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 13

## Conditional Entropy $H(Y|X)$

X = College Major  
Y = Likes "Gladiator"

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

**Definition of Conditional Entropy:**

$H(Y|X)$  = The average specific conditional entropy of  $Y$

= if you choose a record at random what will be the conditional entropy of  $Y$ , conditioned on that row's value of  $X$

= Expected number of bits to transmit  $Y$  if both sides will know the value of  $X$

$$= \sum_j \text{Prob}(X=v_j) H(Y|X=v_j)$$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 14

## الانتروبيا الشرطية المحددة $H(Y|X = v)$

تعريف الانتروبيا الشرطية المحددة:

Y = "Gladiator" إعجاب

X = College Major

ح(ص | س = ت) = إنتروبيا ص من بين تلك السجلات فقط التي X له قيمة/الخامس

مثال:

- ح(ص | س = رياضيات) = 1
- ح(ص | س = التاريخ) = 0
- ح(ص | س = CS) = 0

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 13

حقوق النشر © 2001 ، أندرو ديليو مور

## الانتروبيا الشرطية $H(Y|X)$

تعريف الانتروبيا الشرطية:

Y = "Gladiator" إعجاب

X = College Major

ح(ص | X) = متوسط الانتروبيا الشرطية المحددة لـ ص

= إذا اخترت سجلاً عشوائياً ، فماذا سيكون الانتروبيا الشرطية ص ، مشروطة بقيمة هذا الصف X

= العدد المتوقع من البتات للإرسال ص إذا كان كلا الجانبين سيعرف قيمة X

$\Sigma$  = احتمال  $(X = v)$  ح(ص |  $X = v$ )

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 14

حقوق النشر © 2001 ، أندرو ديليو مور

## Conditional Entropy

X = College Major

Y = Likes "Gladiator"

**Definition of Conditional Entropy:**

$H(Y|X)$  = The average conditional entropy of Y

$$= \sum_j \text{Prob}(X=v_j) H(Y | X = v_j)$$

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

**Example:**

| $v_j$   | $\text{Prob}(X=v_j)$ | $H(Y   X = v_j)$ |
|---------|----------------------|------------------|
| Math    | 0.5                  | 1                |
| History | 0.25                 | 0                |
| CS      | 0.25                 | 0                |

$$H(Y|X) = 0.5 * 1 + 0.25 * 0 + 0.25 * 0 = 0.5$$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 15

## Information Gain

X = College Major

Y = Likes "Gladiator"

**Definition of Information Gain:**

$IG(Y|X)$  = I must transmit Y.  
How many bits on average would it save me if both ends of the line knew X?

$$IG(Y|X) = H(Y) - H(Y | X)$$

**Example:**

- $H(Y) = 1$
- $H(Y | X) = 0.5$
- Thus  $IG(Y|X) = 1 - 0.5 = 0.5$

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 16

## الانتروبيا الشرطية

تعريف الانتروبيا الشرطية:  $Y = \text{"Gladiator" إعجاب}$

$X = \text{College Major}$

ح(ص| $X$ ) = متوسط الانتروبيا الشرطية  
لص

$\Sigma$  = احتمال ( $X = v$ ) ح(ص| $X = v$ )

مثال:

| الخامس  | احتمال ( $X = v$ ) | ح(ص  $X = v$ ) |
|---------|--------------------|----------------|
| رياضيات | 0.5                | 1              |
| تاريخ   | 0.25               | 0              |
| CS      | 0.25               | 0              |

$$0.5 = 0 * 0.25 + 0 * 0.25 + 1 * 0.5 = \text{ح(ص|} X \text{)}$$

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 15

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## كسب المعلومات

تعريف كسب المعلومات:  $Y = \text{"Gladiator" إعجاب}$

$X = \text{College Major}$

$Y (IG | X) =$  يجب أن أنقل ص. كم عدد  
البتات في المتوسط ستوفر لي إذا  
عرف طريقي السطر  $X$ ؟

$$Y (IG | X) = \text{ح(ص)} - \text{ح(ص|} X \text{)}$$

مثال:

- $\text{ح(ص)} = 1$
- $\text{ح(ص | س)} = 0.5$
- وبالتالي  $Y (IG | X) = 1 - 0.5 = 0.5$



| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 16

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور












# Information Gain Example

|                |        |   |      |   |   |
|----------------|--------|---|------|---|---|
| wealth values: |        | poor  | rich |   |   |
| gender         | Female | 14423   | 1769 |  | $H(\text{wealth}   \text{gender} = \text{Female}) = 0.497654$ |
|                | Male   | 22732   | 9918 |  | $H(\text{wealth}   \text{gender} = \text{Male}) = 0.885847$   |
|                |        | $H(\text{wealth}) = 0.793844$                   |      |   | $H(\text{wealth}   \text{gender}) = 0.757154$                 |
|                |        | $IG(\text{wealth}   \text{gender}) = 0.0366896$ |      |   |   |

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 17

# Another example

|                |     |   |      |   |  |
|----------------|-----|---|------|---|--|
| wealth values: |     | poor  | rich |   |  |
| agegroup       | 10s | 2507  | 3    |  | $H(\text{wealth}   \text{agegroup} = 10s) = 0.0133271$ |
|                | 20s | 11262   | 743  |  | $H(\text{wealth}   \text{agegroup} = 20s) = 0.334906$  |
|                | 30s | 9468  | 3461 |  | $H(\text{wealth}   \text{agegroup} = 30s) = 0.838134$  |
|                | 40s | 6738  | 3986 |  | $H(\text{wealth}   \text{agegroup} = 40s) = 0.951961$  |
|                | 50s | 4110  | 2509 |  | $H(\text{wealth}   \text{agegroup} = 50s) = 0.957376$  |
|                | 60s | 2245  | 809  |  | $H(\text{wealth}   \text{agegroup} = 60s) = 0.834049$  |
|                | 70s | 668   | 147  |  | $H(\text{wealth}   \text{agegroup} = 70s) = 0.680882$  |
|                | 80s | 115   | 16   |  | $H(\text{wealth}   \text{agegroup} = 80s) = 0.535474$  |
|                | 90s | 42  | 13   |  | $H(\text{wealth}   \text{agegroup} = 90s) = 0.788941$  |
|                |     | $H(\text{wealth}) = 0.793844$                     |      |   | $H(\text{wealth}   \text{agegroup}) = 0.709463$        |
|                |     | $IG(\text{wealth}   \text{agegroup}) = 0.0843813$ |      |   |  |

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 18

## مثال على كسب المعلومات

wealth values: poor rich

| gender | poor  | rich | H(wealth   gender) |
|--------|-------|------|--------------------|
| Female | 14423 | 1769 | 0.497654           |
| Male   | 22732 | 9918 | 0.885847           |

$H(\text{wealth}) = 0.793844$      $H(\text{wealth}|\text{gender}) = 0.757154$   
 $IG(\text{wealth}|\text{gender}) = 0.0366896$

كسب المعلومات: الشريحة 17

حقوق النشر © 2001 ، أندرو ديليو مور

## مثال آخر

wealth values: poor rich

| agegroup | poor  | rich | H(wealth   agegroup) |
|----------|-------|------|----------------------|
| 10s      | 2507  | 3    | 0.0133271            |
| 20s      | 11262 | 743  | 0.334906             |
| 30s      | 9468  | 3461 | 0.838134             |
| 40s      | 6738  | 3986 | 0.951961             |
| 50s      | 4110  | 2509 | 0.957376             |
| 60s      | 2245  | 809  | 0.834049             |
| 70s      | 668   | 147  | 0.680882             |
| 80s      | 115   | 16   | 0.535474             |
| 90s      | 42    | 13   | 0.788941             |

$H(\text{wealth}) = 0.793844$      $H(\text{wealth}|\text{agegroup}) = 0.709463$   
 $IG(\text{wealth}|\text{agegroup}) = 0.0843813$

كسب المعلومات: الشريحة 18

حقوق النشر © 2001 ، أندرو ديليو مور

## Relative Information Gain

X = College Major

Y = Likes "Gladiator"

**Definition of Relative Information Gain:**

$RIG(Y|X)$  = I must transmit  $Y$ , what fraction of the bits on average would it save me if both ends of the line knew  $X$ ?

| X       | Y   |
|---------|-----|
| Math    | Yes |
| History | No  |
| CS      | Yes |
| Math    | No  |
| Math    | No  |
| CS      | Yes |
| History | No  |
| Math    | Yes |

$$RIG(Y|X) = H(Y) - H(Y|X) / H(Y)$$

**Example:**

- $H(Y|X) = 0.5$
- $H(Y) = 1$
- *Thus*  $IG(Y|X) = (1 - 0.5)/1 = 0.5$

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 19

## What is Information Gain used for?

Suppose you are trying to predict whether someone is going live past 80 years. From historical data you might find...

- $IG(\text{LongLife} | \text{HairColor}) = 0.01$
- $IG(\text{LongLife} | \text{Smoker}) = 0.2$
- $IG(\text{LongLife} | \text{Gender}) = 0.25$
- $IG(\text{LongLife} | \text{LastDigitOfSSN}) = 0.00001$

IG tells you how interesting a 2-d contingency table is going to be.

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 20

## كسب المعلومات النسبية

تعريف كسب المعلومات النسبية:  $Y = \text{"Gladiator" إعجاب}$

$X = \text{College Major}$

$Y(X| \text{RIG}) = \text{يجب أن أنقل نعم، أي جزء من}$   
 البتات في المتوسط سيوفر لي إذا عرف  
 طرفي السطر  $X$ ؟

$$Y(X| \text{RIG}) = \text{ح(ص)} - \text{ح(ص| ح(ص) / H)}$$

مثال:

- $\text{ح(ص)} = 0.5$
- $\text{ح(ص)} = 1$
- وهكذا  $Y(X| \text{IG}) = 0.5 = 1 / (0.5 - 1)$

| ص   | X       |
|-----|---------|
| نعم | رياضيات |
| لا  | تاريخ   |
| نعم | CS      |
| لا  | رياضيات |
| لا  | رياضيات |
| نعم | CS      |
| لا  | تاريخ   |
| نعم | رياضيات |

كسب المعلومات: الشريحة 19

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور

## ما هو مكاسب المعلومات المستخدمة؟

افتراض أنك تحاول توقع ما إذا كان شخص ما سيعيش بعد 80 عاماً. من البيانات التاريخية التي قد تجدها ...

- $IG(\text{لون الشعر}) = 0.01$
- $IG(\text{لونج لايف | مدخن}) = 0.2$
- $IG(\text{الجنس | LongLife}) = 0.25$
- $IG(\text{LongLife | LastDigitOfSSN}) = 0.00001$

يخبرك IG كيف سيكون جدول الطوارئ ثنائي الأبعاد مثيراً للاهتمام.

كسب المعلومات: الشريحة 20

حقوق النشر © 2001 ، 2003 ، أندرو ديليو مور