

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  
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: BAACBADCADDAD...

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 الشريحة.  
سيسعدني إذا وجدت هذه المقدمة مفيدة في إلقاء محاضراتك  
الخاصة. لا تتردد في استخدام هذه الشريحة حرفياً، أو تعدلها  
لتلائم احتياجاتك الخاصة. أصول PowerPoint متوفرة، إذا  
كنت تستخدم جزءاً كبيراً من هذه الشريحة في محاضراتك  
الخاصة، فيرجى تضمين هذه الرسالة، أو الرابط التالي إلى  
مستودع المتصدر لدروس آندره التعلميمية: <http://www.cs.cmu.edu/~awm/tutorials>  
التعليمات والتوصيات الواردة  
<http://www.cs.cmu.edu>  
بامتنان.

# كسب المعلومات

أندرو دبليو مور

أستاذ

كلية علوم الكمبيوتر بجامعة

كارنيجي ميلون

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

awm@cs.cmu.edu

412-268-7599

حقوق النشر © 2001 ، 2003 ، آندرو دبليو مور

## بت

أنت تشاهد مجموعة من العينات العشوائية المستقلة لـ X

ترى 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: ... BAACBADCADDAA

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

... 01000010010011011001111100

Kelvin 2

حقوق النشر © 2001 ، 2003 ، آندرو دبليو مور

## Fewer Bits

Someone tells you that the probabilities are not equal

$$P(X=A) = 1/2 \quad P(X=B) = 1/4 \quad P(X=C) = 1/8 \quad P(X=D) = 1/8$$

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

$$P(X=A) = 1/2 \quad P(X=B) = 1/4 \quad P(X=C) = 1/8 \quad P(X=D) = 1/8$$

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

يُخبرك أحدهم أن الاحتمالات ليست متساوية

$$P(X = D) = 1/8 \quad P(X = C) = 1/8 \quad P(X = B) = 1/4 \quad P(X = A) = 1/2$$

انه ممکن...

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

كتاب المعلومات: الشريحة 3

حقوق النشر © 2003 ، أندرو دبليو مور

## بت أقل

يُخبرك أحدهم أن الاحتمالات ليست متساوية

$$P(X = D) = 1/8 \quad P(X = C) = 1/8 \quad P(X = B) = 1/4 \quad P(X = A) = 1/2$$

انه ممکن...

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)

كتاب المعلومات: الشريحة 4

حقوق النشر © 2003 ، أندرو دبليو مور

## Fewer Bits

Suppose there are three equally likely values...

$$P(X=A) = 1/3 \quad P(X=B) = 1/3 \quad P(X=C) = 1/3$$

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...  $V_1, V_2, \dots, V_m$

$$P(X=V_1) = p_1 \quad P(X=V_2) = p_2 \quad \dots \quad P(X=V_m) = p_m$$

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

$$\begin{aligned} H(X) &= -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_m \log_2 p_m \\ &= -\sum_{j=1}^m p_j \log_2 p_j \end{aligned}$$

$H(X)$  = 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

افترض أن هناك ثلاثة قيم متساوية الاحتمال ...

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

هذا تميز ساذج ، يكلف 2 بت لكل رمز

00	أ
01	ب
10	ج

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.

كتب المعلومات: الشريحة 5

حقوق النشر © 2003 ، 2001 ، أندرو ديليو مور

## الحالة العامة

افترض أن  $X$  يمكن أن يكون له واحد من القيم... الخامس<sub>1</sub> ، الخامس<sub>2</sub> ... الخامس<sub>m</sub>

$$P(X = \text{ص}_1) = P_1, \quad \dots, \quad P(X = \text{ص}_m) = P_m$$

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

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

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

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

- "إنتروبيا عالية" تعني أن  $X$  ناتجة عن توزيع موحد (مملاً)
- يعني "إنتروبيا منخفض" أن  $X$  من توزيع (قمم ووديان) متتنوع

كتب المعلومات: الشريحة 6

حقوق النشر © 2003 ، 2001 ، أندرو ديليو مور

## General Case

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

$$P(X=V_1) = p_1 \quad P(X=V_2) = p_2 \quad \dots \quad P(X=V_m) = p_m$$

What's the smallest possible number of symbol, X's distribution? A histogram of the frequency distribution of values of X would be flat

$H(X)$

$$= - \sum p_j \log_2 p_j$$

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

$H(X) =$  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...  $V_1, V_2, \dots, V_m$

$$P(X=V_1) = p_1 \quad P(X=V_2) = p_2 \quad \dots \quad P(X=V_m) = p_m$$

What's the smallest possible number of symbol, X's distribution? A histogram of the frequency distribution of values of X would be flat

$H(X)$

$$= - \sum p_j \log_2 p_j$$

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

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 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

الحالة العامة

افترض أن  $X$  يمكن أن يكون له واحد من القيم ... الخامس<sub>1</sub> ، الخامس<sub>2</sub> ، ... الخامس<sub>m</sub>

$$\mu_{\text{ص}} = ({}_1P)X = V. \quad \dots \quad \mu_{\text{ص}} = ({}_2P)X = V. \quad \mu_{\text{ص}} = ({}_3P)X = V.$$

$$\sum_{i=1}^n \text{صی سجل}^2 \text{ صی} =$$

سواتر انتروپی

- يعني إنتروروبية منخفض "أن X من توزيع (قلم ووديان) متنوع"
  - يعني إنتروروبية عالية "تعني أن X ناتجة عن توزيع موحد (مم)"

الشرح المعلمات: كبس 7

حقوق النشر © 2003 ، 2001 ، أندرو ديليو مور

الحالة العامة

افتراض أن  $X$  يمكن أن يكون له واحد من القيم... الخامس<sup>1</sup>, الخامس<sup>2</sup>, ... الخامس<sup>m</sup>

$$\begin{array}{l} \text{ص} = ({}_2P)X = V. \\ \dots \\ \text{ص} = ({}_2P)X = V. \\ \text{ص} = ({}_1P)X = V. \end{array}$$

$$\sum_{i=1}^{n-2} \text{ص.ب. المصغرى} = \text{المضار الثانى زس} - \text{اص.ب. القيم الإلكترونية}$$

$X = \text{انتروسا}(X)$

- "إنتروليبا عالية" تعني أن  $X$  ناتجة عن توزيع موحد (مم)
  - يعني "إنتروليبا منخفض" أن  $X$  من توزيع (قمم وواديان) متنوع

كتاب المعلمون في الشريعة

© ٢٠٠٣ ، ٢٠٠١ النشر ، أندونيسيا

## 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

## الانتروبيافي قشرة الجوز



ارتفاع الانتروبيا



إنتروبيا منخفضة

كبس المعلومات: الشريحة 9

حقوق النشر © 2001 ، 2003 ، أندرو دبليو مور

## الانتروبيافي قشرة الجوز



ارتفاع الانتروبيا



إنتروبيا منخفضة

كبس المعلومات: الشريحة 10

حقوق النشر © 2001 ، 2003 ، أندرو دبليو 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

- $P(LikeG = Yes) = 0.5$
- $P(Major = Math \& LikeG = No) = 0.25$
- $P(Major = Math) = 0.5$
- $P(LikeG = Yes | Major = History) = 0$

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 بالخرج  $Y$  ولدي إدخال  $X$

لنفترض أن هذا يعكس الاحتمالات الحقيقية تخصص الكلية =  $X$

$Y$  = لاعيات "Gladiator"

من هذه البيانات نقدر  $EG$

$$P(LikeG) = 0.5 \cdot \text{نعم}$$

$$P(LikeG) = 0.25 \cdot \text{الرياضيات} \wedge \text{الرائد}$$

$$0.5 \cdot \text{الفوسفور} \wedge \text{الرائد} = \text{الرياضيات}$$

$$0 \cdot \text{نعم} = P(LikeG) \mid \text{التخصص} = \text{التاريخ}$$

ملحوظة:

$$H(S) = 1.5$$

$$H(S) = 1$$

ص	X
نعم	رياضيات
لا	تاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	تاريخ
نعم	رياضيات

كتاب المعلومات: الشريحة 11

حقوق النشر © 2003 ، أندرو دبليو مور

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

تعريف الانتروبيا الشرطية المحددة:

$Y$  = إعجاب "Gladiator"

$X$  = College Major

$H(S|S = T) = \text{إنتروبيا} S$  من بين تلك السجلات فقط التي لها قيمة الخامس

ص	X
نعم	رياضيات
لا	تاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	تاريخ
نعم	رياضيات

كتاب المعلومات: الشريحة 12

حقوق النشر © 2003 ، أندرو دبليو 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:

$v$  = إعجاب "Gladiator"

$X$  = College Major

$H(S|S = T) = \text{إنتروبيا ص من بين تلك السجلات فقط التي } X \text{ لها قيمة الخامس}$

مثال:

•  $H(S|S = \text{رياضيات}) = 1$

•  $H(S|S = \text{التاريخ}) = 0$

•  $H(S|S = CS) = 0$

S	X
نعم	رياضيات
لا	التاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	التاريخ
نعم	رياضيات

كتب المعلومات: الشريحة 13

حقوق النشر © 2003 ، أندرو دبليو مور

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

تعريف الانتروبيا الشرطية:

$v$  = إعجاب "Gladiator"

$X$  = College Major

$H(S|X) = \text{متوسط الانتروبيا الشرطية المحددة لـ } S$

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

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

$\sum_i P(X_i) H(S|X_i) = v$

S	X
نعم	رياضيات
لا	التاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	التاريخ
نعم	رياضيات

كتب المعلومات: الشريحة 14

حقوق النشر © 2003 ، أندرو دبليو مور

## Conditional Entropy

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 conditional entropy of Y

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

**Example:**

$v_j$	$Prob(X=v_j)$	$H(Y   X = v_j)$
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"

X	Y
Math	Yes
History	No
CS	Yes
Math	No
Math	No
CS	Yes
History	No
Math	Yes

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

Copyright © 2001, 2003, Andrew W. Moore

Information Gain: Slide 16

## الانتروبيا الشرطية

تعريف الانتروبيا الشرطية:  $Y = \text{Entropy}$

$H(X) = \frac{1}{n} \sum P(x_i) \log P(x_i)$  متوسط الانتروبيا الشرطية

$$\sum_i P(x_i) \log P(x_i)$$

مثال:

$P(X=v)$	$H(X=v)$	$P(X=x_i)$	$H(X=x_i)$
1	0.5	0.5	0.5
0	0.25	0.25	0.25
0	0.25	0.25	0.25

$$H(X) = -\sum P(x_i) \log P(x_i) = -0.5 \log 0.5 - 0.25 \log 0.25 - 0.25 \log 0.25 = 0.918$$

ص	X
نعم	رياضيات
لا	تاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	تاريخ
نعم	رياضيات

كسب المعلومات: الشريحة 15

حقوق النشر © 2003 ، أندرو دبليو مور

## كسب المعلومات

تعريف كسب المعلومات:  $Y = \text{Information Gain}$

$IG(X|Y) = \sum P(x_i) H(Y|x_i) = \sum P(x_i) H(Y) - H(Y|x_i)$  يجب أن أنقل ص. كم عدد البيانات في المتوسط ستتوفر لي إذا عرف طرقي السطر X؟

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

مثال:

- $H(Y) = 1$

- $H(Y|X) = 0.5$

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

ص	X
نعم	رياضيات
لا	تاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	تاريخ
نعم	رياضيات

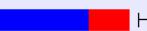
كسب المعلومات: الشريحة 16

حقوق النشر © 2003 ، أندرو دبليو مور

# Information Gain Example

wealth values: poor rich

gender Female 14423 1769  H( wealth | gender = Female ) = 0.497654

Male 22732 9918  H( wealth | gender = 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( wealth | agegroup = 10s ) = 0.0133271

20s 11262 743  H( wealth | agegroup = 20s ) = 0.334906

30s 9468 3461  H( wealth | agegroup = 30s ) = 0.838134

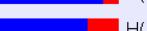
40s 6738 3986  H( wealth | agegroup = 40s ) = 0.951961

50s 4110 2509  H( wealth | agegroup = 50s ) = 0.957376

60s 2245 809  H( wealth | agegroup = 60s ) = 0.834049

70s 668 147  H( wealth | agegroup = 70s ) = 0.680882

80s 115 16  H( wealth | agegroup = 80s ) = 0.535474

90s 42 13  H( wealth | 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	Female	14423	1769	H( wealth   gender = Female ) = 0.497654
	Male	22732	9918	H( wealth   gender = Male ) = 0.885847

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

كسب المعلومات: الشريحة 17

حقوق النشر © 2001 ، 2003 ، أندرو دبليو مور

## مثال آخر

wealth values: poor rich

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

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

كسب المعلومات: الشريحة 18

حقوق النشر © 2001 ، 2003 ، أندرو دبليو 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:  $Y = \text{إعجاب "Gladiator"}$

$X = \text{College Major}$

$X / RIG =$  يجب أن أنقل نعم، أي جزء من  
البيانات في المتوسط سيوفر لي إذا عرف  
طرف السطر  $X$ ؟

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

مثال:

$$\text{ح}(\text{ص}) = 0.5$$

$$\text{ح}(\text{ص}) = 1$$

$$0.5 = 1 / (0.5 - 1) = 1 / (-0.5) = -2$$

ص	X
نعم	رياضيات
لا	تاريخ
نعم	CS
لا	رياضيات
لا	رياضيات
نعم	CS
لا	تاريخ
نعم	رياضيات

Kelvin

حقوق النشر © 2003 ، أندرو دبليو مور

## ما هو مكاسب المعلومات المستخدمة؟

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

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

يخبرك  $IG$  كيف سيكون جدول الطوارئ ثنائي الأبعاد مثيراً للاهتمام.

Kelvin

حقوق النشر © 2003 ، أندرو دبليو مور