Information Theory

Mohamed Hamada

Software Engineering Lab
The University of Aizu

Email: hamada@u-aizu.ac.jp
URL: http://www.u-aizu.ac.jp/~hamada

Today's Topics

- **Entropy**
- **Conditional Entropy**
- **Example**
- **Joint Entropy**
- **Chain Rule**

Entropy H(S)

- Entropy is the average information content of a source

$$H(S) = E[I(s_k)]$$

$$H(S) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

Entropy

Comments

Entropy is a measure of how much information is encoded in a message. Higher the entropy, higher the information content.

We could also say entropy is a measure of uncertainty in a message.

Information and uncertainty are equivalent concepts.

Entropy gives the actual number of bits of information contained in a message source.

Example: if the probability of the character `e` appearing in this slide is 1/16, then the information content of this character is 4 bits.

So the character string `eeeee` has a total of 20 bits (contrast this to using an 8-bit ASCII coding that could result in 40 bits to represent `eeeee`.

Conditional Entropy

Conditional Entropy

H(Y|X=v)

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

Conditional Entropy H(Y|X=v)

Example **If I know the student's major could I predict if he likes computer games?**

Input: X = College Major

Output: Y = Likes "Computer Games"

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

Marginal distribution for X

X \ Y	Yes	No	
Math	0.25	0.25	P(X=Math)=0.5
CS	0.25	0	P(X=CS)=0.25
History	0	0.25	P(X=History)=0.25
	P(Y=Yes)=0.5	P(Y=No)=0.5	1

Marginal distribution for Y

7

Conditional Entropy H(Y|X=v)

Example **If I know the student's major could I predict if he likes computer games?**

Input: X = College Major

Output: Y = Likes "Computer Games"

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

X \ Y	Yes	No	
Math	0.25	0.25	P(X=Math)=0.5
CS	0.25	0	P(X=CS)=0.25
History	0	0.25	P(X=History)=0.25
	P(Y=Yes)=0.5	P(Y=No)=0.5	1

H(X) = - 0.5 log 0.5 – 0.25 log 0.25 – 0.25 log 0.25 =1.5

H(Y) = - 0.5 log 0.5 - 0.5 log 0.5 =1

8

Specific Conditional Entropy H(Y|X=v)

Example

Definition of Specific Conditional Entropy:

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

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

9

Specific Conditional Entropy H(Y|X=v)

Example

Definition of Specific Conditional Entropy:

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

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

Example:

H(Y|X=Math) = - S p(y|X=Math) log p(y|X=Math)
y ?{Yes, No}

=-p(Yes|X=Math) log p(Yes|X=Math) -p(No|X=Math) log p(No|X=Math)

=-0.5 log 0.5 – 0.5 log 0.5 =1

H(Y|X=Math) = 1

10

Specific Conditional Entropy H(Y|X=v)

Example

Definition of Specific Conditional Entropy:

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

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

Example:

H(Y|X=History) =-S p(y|X=History) log p(y|X=History)
y ?{Yes, No}

=-p(Yes|X=History) log p(Yes|X=History) -p(No|X=History) log p(No|X=History)

=-0 log 0 – 1 log 1 =0

H(Y|X=History) = 0

11

Specific Conditional Entropy H(Y|X=v)

Example

Definition of Specific Conditional Entropy:

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

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

Example:

H(Y|X=CS) =-S p(y|X=CS) log p(y|X=CS)
y ?{Yes, No}

=-p(Yes|X=CS) log p(Yes|X=CS) -p(No|X=CS) log p(No|X=CS)

=-1 log 1 – 0 log 0 =0

H(Y|X=CS) = 0

12

Specific Conditional Entropy H(Y|X=v)

Example

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

Conditional Entropy H(Y|X)

Is the amount of information contained in Y such that X is given

Example

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

= S_j Prob(X=v_j) H(Y | X = v_j)

Conditional Entropy

Example

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

= S_jProb(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

Joint Entropy

Joint Entropy

Is the amount of information contained in both events X and Y

$$H(X, Y) = -S_{X,Y}\ p(x,y) \log p(x,y)$$

Joint Entropy

Chain Rule

Relationship between conditional and joint entropy

$$H(X,Y) = H(X) + H(Y|X)$$

Proof:

$$H(X,Y) = -\sum_{X,Y} P(x,y)\log_2 P(x,y) = -\sum_{X,Y} P(x,y)\log_2 P(x)P(y|x)$$

$$= -\sum_{X} P(x)\sum_{Y} P(y|x)\log_2 P(x) - \sum_{X,Y} P(x,y)\log_2 P(y|x) = H(X) + H(Y|X)$$

$$H(X|Y) = -\sum_{X,Y} P(x,y)\log_2 P(x|y)$$

Joint Entropy

Is the amount of information contained in both events X and Y

H(X,Y) = H(X) + H(Y|X)

Also H(X,Y) = H(Y) + H(X|Y)

- Intuition: first describe Y and then X given Y
- From this: H(X) – H(X|Y) = H(Y) – H(Y|X)

19

Joint Entropy

Example

H(X,Y) = H(X) + H(Y|X)

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

v_j	Prob(X=v_j)	H(Y | X = v_j)
Math	0.5	1
History	0.25	0
CS	0.25	0

H(X) =1.5

H(Y) =1

H(Y|X) = 0.5

H(X,Y) = H(X) + H(Y|X)=1.5+0.5=2

20

Joint Entropy

Comments

H(X,Y) = H(X) + H(Y|X)
= H(Y) + H(X|Y)

H(X,Y,Z) = H(X) + H(Y|X) + H(Z|XY)
£ H(X) + H(Y) + H(Z)

21