X Y

— —

case (1)

simple example:

$p(x,y)$	o	y	i
n	1/4	0	0
m	0	1/4	0
b	0	1/4	1/4

marginals

$P(X = n) = 1/4$

$P(X = m) = 1/4$

$P(X = b) = 1/2$

$P(Y = o) = 1/4$

$P(Y = y) = 1/2$

$P(Y = i) = 1/4$

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$-\sum_x p(x) \cdot \log(p(x)) \quad + \sum_{xy} p(x,y) \log(p(x|y))$$

$$H(x) - \sum_y p(y) \cdot H(X|Y=y)$$

averages across the entire distribution and not just the case of Y = "o"

Mutual information is symmetric "on average" reduction in uncertainty

is now equivalent to a class problem and thus not solving ahead