

Takeaway notes:

$$I(A, B) = H(A) + H(B) - H(A, B)$$

↳ Amount of information gained by knowing the value of one of A & B

$$VI(A, B) = \underbrace{[H(A, B) - H(A)]}_{H(B|A)} + \underbrace{[H(A, B) - H(B)]}_{H(A|B)}$$

Each cond term is the amount of information needed to encode the value of 1 var after knowing the other.

This is the best formulation for implementation.

