Information and Coding Theory

MA41024/ MA60020/ MA60262

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 5
January 17, 2023

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

First observe that it is enough to prove it for natural numbers a_i, b_i.

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

First observe that it is enough to prove it for natural numbers a_i, b_i.

- ▷ Consider disjoint subsets of natural numbers $A_1, \ldots, A_n$ such that $|A_i| = a_i$

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

First observe that it is enough to prove it for natural numbers a_i, b_i.

- ▷ Consider disjoint subsets of natural numbers $A_1, \ldots, A_n$ such that $|A_i| = a_i$
- ▷ do the same, $|B_i| = b_i$

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

An information theoretic proof of the Cauchy-Schwarz inequality[1]

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

First observe that it is enough to prove it for natural numbers a_i, b_i.

- Consider disjoint subsets of natural numbers $A_1, \ldots, A_n$ such that $|A_i| = a_i$
- do the same, $|B_i| = b_i$
- Consider rectangles in the xy plane with x coordinates in the sets A_i and y coordinates in the sets B_i.

---

[1] Ehud Friedgut, Hypergraphs, entropy, and inequalities, The American Mathematical Monthly 111 (2004), no. 9, 749–760. 6, 8

Entropy

- Pick two points (X_1, Y_1) and (X_2, Y_2) at random

Entropy

- Pick two points (X_1, Y_1) and (X_2, Y_2) at random
  - Pick a rectangle R with probability proportional to its area i.e.

$$\mathbb{P}[R = r_i] = \frac{a_i \cdot b_i}{\sum_j a_j \cdot b_j}$$

    where $r_i = A_i \times B_i$ is the rectangle with $a_i \cdot b_i$ points

Entropy

- Pick two points (X_1, Y_1) and (X_2, Y_2) at random
  - Pick a rectangle R with probability proportional to its area i.e.

$$\mathbb{P}[R = r_i] = \frac{a_i \cdot b_i}{\sum_j a_j \cdot b_j}$$

    where $r_i = A_i \times B_i$ is the rectangle with $a_i \cdot b_i$ points
  - Pick two points (X_1, Y_1) and (X_2, Y_2) independently from R (uniform distribution)

Entropy

- Pick two points (X_1, Y_1) and (X_2, Y_2) at random
  - Pick a rectangle R with probability proportional to its area i.e.

$$\mathbb{P}[R = r_i] = \frac{a_i \cdot b_i}{\sum_j a_j \cdot b_j}$$

    where $r_i = A_i \times B_i$ is the rectangle with $a_i \cdot b_i$ points
  - Pick two points (X_1, Y_1) and (X_2, Y_2) independently from R (uniform distribution)
- since the sets are disjoint, specifying any of the variables X_1, Y_1, X_2, Y_2 reveals the rectangle R i.e. $H(X_1, Y_1, R) = H(X_2, Y_2, R) = H(X_1, Y_1) = H(X_2, Y_2) \stackrel{\text{how}}{=} \log\left(\sum_i a_i \cdot b_i\right)$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$\begin{aligned} & H(X_1, Y_1, R) + H(X_2, Y_2, R) \\ = \;& 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \end{aligned}$$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$
\begin{aligned}
& H(X_1, Y_1, R) + H(X_2, Y_2, R) \\
= \; & 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \\
= \; & 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R)
\end{aligned}
$$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$\begin{aligned} & H(X_1, Y_1, R) + H(X_2, Y_2, R) \\ = \; & 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \\ = \; & 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R) \\ = \; & 2H(R) + H(X_1, X_2|R) + H(Y_1, Y_2|R) \end{aligned}$$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$\begin{aligned}
& H(X_1, Y_1, R) + H(X_2, Y_2, R) \\
= \; & 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \\
= \; & 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R) \\
= \; & 2H(R) + H(X_1, X_2|R) + H(Y_1, Y_2|R) \\
= \; & H(X_1, X_2, R) + H(Y_1, Y_2, R)
\end{aligned}$$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$\begin{aligned}
& H(X_1, Y_1, R) + H(X_2, Y_2, R) \\
= \; & 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \\
= \; & 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R) \\
= \; & 2H(R) + H(X_1, X_2|R) + H(Y_1, Y_2|R) \\
= \; & H(X_1, X_2, R) + H(Y_1, Y_2, R) \\
= \; & H(X_1, X_2) + H(Y_1, Y_2)
\end{aligned}$$

Entropy

Since given R, X_1, X_2, Y_1, Y_2 are independent, we have

$$\begin{aligned}
& H(X_1, Y_1, R) + H(X_2, Y_2, R) \\
= \; & 2H(R) + H(X_1, Y_1|R) + H(X_2, Y_2|R) \\
= \; & 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R) \\
= \; & 2H(R) + H(X_1, X_2|R) + H(Y_1, Y_2|R) \\
= \; & H(X_1, X_2, R) + H(Y_1, Y_2, R) \\
= \; & H(X_1, X_2) + H(Y_1, Y_2) \\
\leq \; & \log\left(\sum_i a_i^2\right) + \log\left(\sum_i b_i^2\right)
\end{aligned}$$

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?

Example X represents the roll of a fair 6-sided die, and Y represents whether the roll is even (0 if even, 1 if odd)

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?
Example X represents the roll of a fair 6-sided die, and Y represents whether the roll is even (0 if even, 1 if odd)

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

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?
Example X represents the roll of a fair 6-sided die, and Y represents whether the roll is even (0 if even, 1 if odd)

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

Then (expanding the formula of entropy)

$\rightarrow\ I(X;Y) \geq 0$

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?
Example X represents the roll of a fair 6-sided die, and Y represents whether the roll is even (0 if even, 1 if odd)

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

Then (expanding the formula of entropy)

$\rightarrow$ $I(X;Y) \geq 0$

$\rightarrow$ $I(X;Y) = I(Y;X)$

Mutual information

The mutual information (MI) between two random variables X and Y is defined as

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

Question What is the difference between correlation and MI?
Example X represents the roll of a fair 6-sided die, and Y represents whether the roll is even (0 if even, 1 if odd)

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

Then (expanding the formula of entropy)

$\rightarrow$ $I(X;Y) \geq 0$

$\rightarrow$ $I(X;Y) = I(Y;X)$

Homework Let X, Y be two variables with $X \vee Y = 1$, $X \in \{0,1\}$, $Y \in \{0,1\}$ such that $(X,Y) = (1,0)$, $(X,Y) = (0,1)$ and $(X,Y) = (1,1)$ with probabilities $1/3$. Then calculate $I(X;Y)$

Mutual information

Conditional mutual information

$$I(X; Y|Z) = \mathbb{E}_Z[I(X|Z = z; Y|Z = z)]$$

Mutual information

Conditional mutual information

$$
\begin{aligned}
I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z;Y|Z=z)] \\
&= \mathbb{E}_Z[H(X|Z=z) - H(X|Y,Z=z)]
\end{aligned}
$$

Mutual information

Conditional mutual information

$$\begin{aligned} I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z;Y|Z=z)] \\ &= \mathbb{E}_Z[H(X|Z=z) - H(X|Y,Z=z)] \\ &= H(X|Z) - H(X|Y,Z) \end{aligned}$$

Mutual information

Conditional mutual information

$$\begin{aligned} I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z;Y|Z=z)] \\ &= \mathbb{E}_Z[H(X|Z=z) - H(X|Y,Z=z)] \\ &= H(X|Z) - H(X|Y,Z) \end{aligned}$$

Example Let (X, Y, Z) be a random variable with $Z = X \oplus Y$, $X \in \{0, 1\}$, $Y \in \{0, 1\}$ such that $(X, Y, Z) = (x, y, z)$ are equally likely. Then check that $I(X; Y) = 0$ and

$$I(X;Y|Z) = \mathbb{E}_Z[I(X|Z=z);Y|Z=z]$$

Mutual information

Conditional mutual information

$$\begin{aligned} I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z;Y|Z=z)] \\ &= \mathbb{E}_Z[H(X|Z=z) - H(X|Y,Z=z)] \\ &= H(X|Z) - H(X|Y,Z) \end{aligned}$$

Example Let (X, Y, Z) be a random variable with $Z = X \oplus Y$, $X \in \{0, 1\}$, $Y \in \{0, 1\}$ such that $(X, Y, Z) = (x, y, z)$ are equally likely. Then check that $I(X;Y) = 0$ and

$$\begin{aligned} I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z);Y|Z=z] \\ &= \frac{1}{2}I(X|Z=0;Y|Z=0) + \frac{1}{2}I(X|Z=1;Y|Z=1) \end{aligned}$$

Mutual information

Conditional mutual information

$$
\begin{aligned}
I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z;Y|Z=z)] \\
&= \mathbb{E}_Z[H(X|Z=z) - H(X|Y,Z=z)] \\
&= H(X|Z) - H(X|Y,Z)
\end{aligned}
$$

Example Let (X,Y,Z) be a random variable with $Z = X \oplus Y$, $X \in \{0,1\}$, $Y \in \{0,1\}$ such that $(X,Y,Z) = (x,y,z)$ are equally likely. Then check that $I(X;Y) = 0$ and

$$
\begin{aligned}
I(X;Y|Z) &= \mathbb{E}_Z[I(X|Z=z);Y|Z=z] \\
&= \frac{1}{2} I(X|Z=0;Y|Z=0) + \frac{1}{2} I(X|Z=1;Y|Z=1) \\
&= \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1
\end{aligned}
$$

Mutual information

Question What is the conclusion from the above example?

Mutual information

Question What is the conclusion from the above example?

Chain rule of MI: $I((X_1, \ldots, X_m); Y) = \sum_{i=1}^{m} I(X_i; Y | X_1, \ldots, X_{i-1})$

Mutual information

Question What is the conclusion from the above example?

Chain rule of MI: $I((X_1, \ldots, X_m); Y) = \sum_{i=1}^{m} I(X_i; Y | X_1, \ldots, X_{i-1})$

Proof

$$
\begin{aligned}
& I((X_1, \ldots, X_m); Y) \\
= \; & H(X_1, \ldots, X_m) - H(X_1, \ldots, X_m | Y)
\end{aligned}
$$

Mutual information

Question What is the conclusion from the above example?

Chain rule of MI: $I((X_1,\ldots,X_m);Y)=\sum_{i=1}^{m} I(X_i;Y|X_1,\ldots,X_{i-1})$

Proof

$$
\begin{aligned}
& I((X_1,\ldots,X_m);Y) \\
= \; & H(X_1,\ldots,X_m) - H(X_1,\ldots,X_m|Y) \\
= \; & \sum_{i=1}^{m} H(X_i|X_1,\ldots,X_{i-1}) - \sum_{i=1}^{m} H(X_i|Y,X_1,\ldots,X_{i-1})
\end{aligned}
$$

Mutual information

Question What is the conclusion from the above example?

Chain rule of MI: $I((X_1, \ldots, X_m); Y) = \sum_{i=1}^{m} I(X_i; Y|X_1, \ldots, X_{i-1})$

Proof

$$
\begin{aligned}
& I((X_1, \ldots, X_m); Y) \\
= \;& H(X_1, \ldots, X_m) - H(X_1, \ldots, X_m|Y) \\
= \;& \sum_{i=1}^{m} H(X_i|X_1, \ldots, X_{i-1}) - \sum_{i=1}^{m} H(X_i|Y, X_1, \ldots, X_{i-1}) \\
= \;& \sum_{i=1}^{m} [H(X_i|X_1, \ldots, X_{i-1}) - H(X_i|Y, X_1, \ldots, X_{i-1})]
\end{aligned}
$$

Mutual information

Question What is the conclusion from the above example?

Chain rule of MI: $I((X_1,\ldots,X_m);Y)=\sum_{i=1}^{m} I(X_i;Y|X_1,\ldots,X_{i-1})$

Proof

$$
\begin{aligned}
& I((X_1,\ldots,X_m);Y) \\
= \ & H(X_1,\ldots,X_m)-H(X_1,\ldots,X_m|Y) \\
= \ & \sum_{i=1}^{m} H(X_i|X_1,\ldots,X_{i-1})-\sum_{i=1}^{m} H(X_i|Y,X_1,\ldots,X_{i-1}) \\
= \ & \sum_{i=1}^{m} [H(X_i|X_1,\ldots,X_{i-1})-H(X_i|Y,X_1,\ldots,X_{i-1})] \\
= \ & \sum_{i=1}^{m} I(X_i;Y|X_1,\ldots,X_{i-1})
\end{aligned}
$$

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X; Y) \geq I(X; Z)$.

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X; Y) \geq I(X; Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X; Y) \geq I(X; Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

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

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X;Y) \geq I(X;Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - H(X|Y, g(Y)) \\ &\geq H(X) - H(X|g(Y)) = I(X; g(Y)) \end{aligned}$$

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X;Y) \geq I(X;Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - H(X|Y, g(Y)) \\ &\geq H(X) - H(X|g(Y)) = I(X; g(Y)) \end{aligned}$$

From the first line, $I(X;Y) = I(X;(Y, g(Y))) = I(X;(Y, Z))$

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X;Y) \geq I(X;Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - H(X|Y, g(Y)) \\ &\geq H(X) - H(X|g(Y)) = I(X; g(Y)) \end{aligned}$$

From the first line, $I(X;Y) = I(X;(Y, g(Y))) = I(X;(Y,Z))$ However, in general,

$$I(X;(Y,Z)) = I(X;Y) + I(X;Z|Y) = I(X;Y)$$

Mutual information

Markov chain (a memoryless process) An ordered tuple of random variables (X, Y, Z) is said to form a Markov chain if X and Z are independent conditioned on Y. In that case we write as $X \to Y \to Z$.

Question If $X \to Y \to Z$ then $Z \to Y \to X$?

Lemma Data Processing Inequality: Let $X \to Y \to Z$ be a Markov chain. Then $I(X; Y) \geq I(X; Z)$.

Proof Let $Z = g(Y)$ for some g then obviously $X \to Y \to g(Y)$.

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(X) - H(X|Y, g(Y)) \\ &\geq H(X) - H(X|g(Y)) = I(X; g(Y)) \end{aligned}$$

From the first line, $I(X; Y) = I(X; (Y, g(Y))) = I(X; (Y, Z))$ However, in general,

$$I(X; (Y, Z)) = I(X; Y) + I(X; Z|Y) = I(X; Y)$$

Thus,

$$I(X; Y) = I(X; (Y, Z)) = H(X) - H(X|Y, Z) \geq H(X) - H(X|Z) = I(X; Z)$$

Mutual information

Question What is the context of this inequality in digital communication systems?

Mutual information

Question What is the context of this inequality in digital communication systems?

Sufficient statistic For rvs X and Y, a function $g(Y)$ is called a sufficient statistic of Y for X if $I(X;Y) = I(X;g(Y))$ i.e. $g(Y)$ contains all the relevant information about X

Mutual information

Question What is the context of this inequality in digital communication systems?

Sufficient statistic For rvs X and Y, a function $g(Y)$ is called a sufficient statistic of Y for X if $I(X;Y) = I(X;g(Y))$ i.e. $g(Y)$ contains all the relevant information about X

Example Let X take two values x_1 and x_2 with equal probability. Suppose Y is a sequence of n coin tosses with probability of heads given by X. Let $g(Y)$ be the number of heads in Y. Then show that
$I(X;Y) = I(X;g(Y))$.