An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

July 18, 2016

Lecture #2B: Problem solving session-I

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X = x) < H(Y)$

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X=x) < H(Y)$
  ii) $H(Y|X=x) > H(Y)$

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X=x) < H(Y)$
  ii) $H(Y|X=x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X = x) < H(Y)$
  ii) $H(Y|X = x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X = x) < H(Y)$
  ii) $H(Y|X = x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.
- Note that $P_{Y/X}(0/1) = 1$ so that $H(Y/X = 1) = 0$.

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  - i) $H(Y|X=x) < H(Y)$
  - ii) $H(Y|X=x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.
- Note that $P_{Y/X}(0/1) = 1$ so that $H(Y/X = 1) = 0$.
- Similarly, we have $P_{Y/X}(0/0) = 1/2$ so that

$$H(Y/X = 0) = h(1/2) = 1 \text{ bit}$$

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  - i) $H(Y|X=x) < H(Y)$
  - ii) $H(Y|X=x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.
- Note that $P_{Y/X}(0/1) = 1$ so that $H(Y/X = 1) = 0$.
- Similarly, we have $P_{Y/X}(0/0) = 1/2$ so that

$$H(Y/X = 0) = h(1/2) = 1 \text{ bit}$$

- Since $P_Y(1) = 1/4$, we have $H(Y) = h(1/4) = 0.811$ bits. Thus we have

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X=x) < H(Y)$
  ii) $H(Y|X=x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.
- Note that $P_{Y/X}(0/1) = 1$ so that $H(Y/X = 1) = 0$.
- Similarly, we have $P_{Y/X}(0/0) = 1/2$ so that

$$H(Y/X = 0) = h(1/2) = 1 \text{ bit}$$

- Since $P_Y(1) = 1/4$, we have $H(Y) = h(1/4) = 0.811$ bits. Thus we have
  i) $H(Y|X = 1) < H(Y)$

Conditional Entropy

- **Problem # 1:** Give examples of joint random variable X and Y such that
  i) $H(Y|X=x) < H(Y)$
  ii) $H(Y|X=x) > H(Y)$
- **Solutions:** Suppose that the random vector [X,Y,Z] is equally likely to take any of the following four values: [0,0,0],[0,1,0],[1,0,0] and [1,0,1].
- Then $P_X(0) = P_X(1) = 1/2$ so that $H(X) = H(1/2) = 1$ bit.
- Note that $P_{Y/X}(0/1) = 1$ so that $H(Y/X = 1) = 0$.
- Similarly, we have $P_{Y/X}(0/0) = 1/2$ so that

$$H(Y/X = 0) = h(1/2) = 1 \text{ bit}$$

- Since $P_Y(1) = 1/4$, we have $H(Y) = h(1/4) = 0.811$ bits. Thus we have
  i) $H(Y|X = 1) < H(Y)$
  ii) $H(Y|X = 0) > H(Y)$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  i) $I(X;Y|Z) < I(X;Y)$

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  i) $I(X;Y|Z) < I(X;Y)$
  ii) $I(X;Y|Z) > I(X;Y)$

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  i) $I(X;Y|Z) < I(X;Y)$
  ii) $I(X;Y|Z) > I(X;Y)$

i) **Solutions:** Let X, Y and Z form a Markov Chain.

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  i) $I(X;Y|Z) < I(X;Y)$
  ii) $I(X;Y|Z) > I(X;Y)$

i) **Solutions:** Let X, Y and Z form a Markov Chain.

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

- We note that $I(X;Z|Y) = 0$, by Markovity, and $I(X;Z) \geq 0$. Thus,

$$I(X;Y|Z) \leq I(X;Y) \tag{1}$$

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  i) $I(X;Y|Z) < I(X;Y)$
  ii) $I(X;Y|Z) > I(X;Y)$

i) **Solutions:** Let X, Y and Z form a Markov Chain.

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

- We note that $I(X;Z|Y) = 0$, by Markovity, and $I(X;Z) \geq 0$. Thus,

$$I(X;Y|Z) \leq I(X;Y) \tag{1}$$

ii) Let X and Y be independent fair binary random variables, and let $Z = X + Y$.

Mutual Information

- **Problem # 2:** Give examples of joint random variable X, Y and Z such that
  - i) $I(X;Y|Z) < I(X;Y)$
  - ii) $I(X;Y|Z) > I(X;Y)$

i) **Solutions:** Let X, Y and Z form a Markov Chain.

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned}$$

- We note that $I(X;Z|Y) = 0$, by Markovity, and $I(X;Z) \geq 0$. Thus,

$$I(X;Y|Z) \leq I(X;Y) \quad (1)$$

ii) Let X and Y be independent fair binary random variables, and let $Z = X + Y$.

- Then $I(X;Y) = 0$, but $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z) = P(Z=1)H(X|Z=1) = \frac{1}{2}$bit.

Divergence

- **Problem # 3:** Let $P_X(X=0) = P_X(X=1) = 0.5$, $Q_X(X=0) = 0.25, Q_X(X=1) = 0.75$ and $R_X(X=0) = 0.2, R_X(X=1) = 0.8$. Show that triangle inequality does hold for divergence, i.e. $D(P_X||R_X) > D(P_X||Q_X) + D(Q_X||R_X)$

Divergence

- **Problem # 3:** Let $P_X(X=0) = P_X(X=1) = 0.5$, $Q_X(X=0) = 0.25, Q_X(X=1) = 0.75$ and $R_X(X=0) = 0.2, R_X(X=1) = 0.8$. Show that triangle inequality does hold for divergence, i.e. $D(P_X||R_X) > D(P_X||Q_X) + D(Q_X||R_X)$
- **Solution:**

$$\begin{aligned} D(P_X||Q_X) &= 0.5\log\frac{0.5}{0.25} + 0.5\log\frac{0.5}{0.75} = 0.208 \\ D(Q_X||R_X) &= 0.25\log\frac{0.25}{0.2} + 0.75\log\frac{0.75}{0.8} = 0.011 \\ D(P_X||R_X) &= 0.5\log\frac{0.5}{0.2} + 0.5\log\frac{0.5}{0.8} = 0.322 \end{aligned}$$

Divergence

- **Problem # 3:** Let $P_X(X=0) = P_X(X=1) = 0.5$, $Q_X(X=0) = 0.25, Q_X(X=1) = 0.75$ and $R_X(X=0) = 0.2, R_X(X=1) = 0.8$. Show that triangle inequality does hold for divergence, i.e. $D(P_X||R_X) > D(P_X||Q_X) + D(Q_X||R_X)$
- **Solution:**

$$\begin{aligned} D(P_X||Q_X) &= 0.5\log\frac{0.5}{0.25} + 0.5\log\frac{0.5}{0.75} = 0.208 \\ D(Q_X||R_X) &= 0.25\log\frac{0.25}{0.2} + 0.75\log\frac{0.75}{0.8} = 0.011 \\ D(P_X||R_X) &= 0.5\log\frac{0.5}{0.2} + 0.5\log\frac{0.5}{0.8} = 0.322 \end{aligned}$$

- Since, $0.322 > 0.208 + 0.011 = 0.219$, triangular inequality is not satisfied.

Mutual Information

- **Problem # 4:** Consider a discrete memoryless channel with inputs X and outputs Y . The input X takes values from a ternary set with equal probability and it is known that the probability of error for the system is p. Using Fano's lemma, find a lower bound to the mutual information $I(X;Y)$ as a function of p.

Mutual Information

- **Problem # 4:** Consider a discrete memoryless channel with inputs X and outputs Y . The input X takes values from a ternary set with equal probability and it is known that the probability of error for the system is p. Using Fano's lemma, find a lower bound to the mutual information $I(X;Y)$ as a function of p.

- **Solutions:** Mutual information can be written as

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

Mutual Information

- **Problem # 4:** Consider a discrete memoryless channel with inputs X and outputs Y. The input X takes values from a ternary set with equal probability and it is known that the probability of error for the system is p. Using Fano's lemma, find a lower bound to the mutual information $I(X;Y)$ as a function of p.
- **Solutions:** Mutual information can be written as

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

- By Fano's inequality, we get

$$H(X|Y) \leq H(P_e) + P_e \log(3-1) = H(p) + p$$

Mutual Information

- **Problem # 4:** Consider a discrete memoryless channel with inputs X and outputs Y. The input X takes values from a ternary set with equal probability and it is known that the probability of error for the system is p. Using Fano's lemma, find a lower bound to the mutual information $I(X;Y)$ as a function of p.
- **Solutions:** Mutual information can be written as

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

- By Fano's inequality, we get

$$H(X|Y) \leq H(P_e) + P_e \log(3-1) = H(p) + p$$

- Thus

$$I(X;Y) \geq H(X) - H(p) - p = \log 3 - H(p) - p$$

Concave Function

- **Problem # 5:** Let $(X, Y) \sim p(x, y) = p(x)p(y|x)$. the mutual information $I(X; Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$

Concave Function

- **Problem # 5:** Let $(X, Y) \sim p(x, y) = p(x)p(y|x)$. the mutual information $I(X; Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$
- **Solutions:** To prove, we expand the mutual information

$$I(X; Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x)$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Concave Function

- **Problem # 5:** Let $(X, Y) \sim p(x, y) = p(x)p(y|x)$. the mutual information $I(X; Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$
- **Solutions:** To prove, we expand the mutual information

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x)$$

- If $p(y|x)$ is fixed, then $p(y)$ is a linear function of $p(x)$.

Concave Function

- **Problem # 5:** Let $(X, Y) \sim p(x, y) = p(x)p(y|x)$. the mutual information $I(X; Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$
- **Solutions:** To prove, we expand the mutual information

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x)$$

- If $p(y|x)$ is fixed, then $p(y)$ is a linear function of $p(x)$.
- Hence $H(Y)$, which is a concave function of $p(y)$, is a concave function of $p(x)$.

Concave Function

- **Problem # 5:** Let $(X, Y) \sim p(x, y) = p(x)p(y|x)$. the mutual information $I(X; Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$
- **Solutions:** To prove, we expand the mutual information

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x)$$

- If $p(y|x)$ is fixed, then $p(y)$ is a linear function of $p(x)$.
- Hence $H(Y)$, which is a concave function of $p(y)$, is a concave function of $p(x)$.
- The second term is a linear function of $p(x)$. Hence, the difference is a concave function of $p(x)$.