ma4261_hw1

Nguyen Ngoc Khanh - A0275047B

August 2024

1 Problem 1

Let $[m] = \{1, 2, ..., m\}$. A submodular function f is a set-valued function $f : 2^{[m]} \to \mathbb{R}$ satisfying

$$f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{a\}) - f(\mathcal{T})$$

for all $\mathcal{S} \subseteq \mathcal{T}$ and $a \notin \mathcal{T}$. Let $\{X_1, X_2, ..., X_m\}$ be a collection of random variables and $X_{\mathcal{S}} = \{X_i : i \in \mathcal{S}\}$. Thus, submodular functions represent diminishing returns.

1. Show that the set-valued function
$$f(\mathcal{S}) = H(X_{\mathcal{S}})$$
is submodular.

2. Discuss some submodularity properties of mutual information.

Proof.

1. Entropy is submodular
Given $\mathcal{S} \subset \mathcal{T} \subset [m]$ (the case when $\mathcal{S} = \mathcal{T}$ is trivial). Let $X = X_{\mathcal{S}}$, $X \coprod Z = X_{\mathcal{T}}$ and Y is any random variable not in $X \coprod Z$. We will prove that
$$H(X, Y) - H(X) \geq H(X, Z, Y) - H(X, Z)$$
By chain rule, that is equivalent to
$$H(Y|X) \geq H(Y|X, Z)$$
Proposition 1 (conditioning does not increase entropy)**.** *Let X, Y, Z be random variables, then*
$$H(Y|X) \geq H(Y|X, Z)$$
Proof of Proposition 1. We have
$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathfrak{X}} p(X = x) H\left(\frac{Y}{X = x}\right) && \text{(definition of conditional entropy)} \\ &\geq \sum_{x \in \mathfrak{X}} p(X = x) H\left(\frac{Y|Z}{X = x}\right) && \text{(conditioning does not increase entropy)} \\ &= \sum_{x \in \mathfrak{X}} p(X = x) \sum_{z \in \mathfrak{Z}} p(Z = z|X = x) H\left(\frac{Y}{X = x, Z = z}\right) && \text{(definition of conditional entropy)} \\ &= \sum_{x \in \mathfrak{X}} \sum_{z \in \mathfrak{Z}} p(X = x, Z = z) H\left(\frac{Y}{X = x, Z = z}\right) && \text{(definition of conditional probability)} \\ &= H(Y|X, Z) && \text{(definition of conditional entropy)} \end{aligned}$$
where $p(E)$ denotes the probability of event E, $H\left(\frac{X}{E}\right)$ denotes the entropy of variable X given event E occurred, $H\left(\frac{X|Y}{E}\right)$ denotes the conditional entropy of variable X relative to variable Y given event E occurred (entropy of $p(X|E)$ relative to $p(Y|E)$) $\square$

2. Mutual information is submodular in each variable with some assumptions

   Let T be a random variable, define $g : 2^{[m]} \to \mathbb{R}$ by

$$g(\mathcal{S}) = I(X_S; T)$$

   Then, g is submodular. Given $\mathcal{S} \subset \mathcal{T} \subset [m]$. Let $X = X_{\mathcal{S}}, X \coprod Z = X_{\mathcal{T}}$ and Y is any random variable not in $X \coprod Z$. We will prove that

$$I(X, Y; T) - I(X; T) \geq I(X, Z, Y; T) - I(X, Z; T)$$

   Assumption 1. *Suppose that $\{X_i\}$ are independent given T . That is, $H(X_i|T) = H(X_i|TX_j)$*

   We have

$$\begin{aligned} LHS &= I(X, Y; T) - I(X; T) \\ &= H(X, Y) - H(X, Y|T) - H(X) + H(X|T) \\ &= H(Y|X) - H(Y|T, X) \\ &= H(Y|X) - H(Y|T) && \text{(by Assumption 1)} \end{aligned}$$

$$\begin{aligned} RHS &= I(X, Z, Y; T) - I(X, Z; T) \\ &= H(X, Z, Y) - H(X, Z, Y|T) - H(X, Z) + H(X, Z|T) \\ &= H(Y|X, Z) - H(Y|T, X, Z) \\ &= H(Y|X, Z) - H(Y|T) && \text{(by Assumption 1)} \end{aligned}$$

   $LHS \geq RHS$ is then a consequence of Proposition 1

□

2 Problem 2

Denote $X^n = \{X_1, X_2, ..., X_n\}$ and $X^n - X_i = \{X_1, ..., X_{i-1}, X_{i+1}, ..., X_n\}$. Prove that for any $n \geq 2$,

$$H(X^n) \geq \sum_{i=1}^{n} H(X_i|X^n - X_i)$$

Proof.

$$\begin{aligned} H(X^n) &= \sum_{i=1}^{n} H(X_i|X^{i-1}) && \text{(chain rule for entropy)} \\ &\geq \sum_{i=1}^{n} H(X_i|X^n - X_i) && \text{(by Proposition 1: } X^n - X_i \supset X^{i-1}) \end{aligned}$$

□

3 Problem 3

Prove that

$$H(X, Y, Z) \leq \frac{1}{2}(H(X, Y) + H(Y, Z) + H(Z, X))$$

and generalize to the case where there are n random variables

Proof. We will prove the general case when there are n random variables. Denote $X^n = \{X_1, ..., X^n\}$ and $X^n - X_i = \{X_1, ..., X_{i-1}, X_{i+1}, ..., X_n\}$

We have

$$\begin{aligned}
H(X^n) &\geq \sum_{i=1}^{n} H(X_i|X^n - X_i) && \text{(Problem 2)} \\
&= \sum_{i=1}^{n} (H(X^n) - H(X^n - X_i)) && \text{(chain rule for entropy)} \\
&= nH(X^n) - \sum_{i=1}^{n} H(X^n - X_i)
\end{aligned}$$

Then,

$$(n-1)H(X^n) \leq \sum_{i=1}^{n} H(X^n - X_i)$$

When $n = 3$, we have

$$H(X, Y, Z) \leq \frac{1}{2}(H(X, Y) + H(Y, Z) + H(Z, X))$$

□

4 Problem 4

Fano's inequality for list decoding: Recall the proof of Fano's inequality. Now develop a generalization of Fano's inequality for list decoding. Let $(X, Y) \sim P_{XY}$ and let $\mathcal{L}(Y) \subset \hat{\mathcal{X}}$ be a set of size $L \geq 1$ (compare this to an estimator $\hat{X}(Y) \in \mathcal{X}$ which is a set of size $L = 1$). Lower bound the probability of error $Pr(X \notin \mathcal{L}(Y))$ in terms of L, $H(X|\mathcal{L}(Y))$ and $|\mathcal{X}|$

Proof.

Let $\mathcal{L} = \mathcal{L}(Y)$ denote the function on Y that outputs a subset of size L of $\mathcal{X}$, then $\mathcal{L}$ is a random variable. Let $p_e = Pr(X \notin \mathcal{L})$. Let E be a random variable defined by

$$E = 1_{X \notin \mathcal{L}}$$

1. Bound $H(E|\mathcal{L})$

   By Conditioning does not increase entropy

   $$H(E|\mathcal{L}) \leq H(E) = H_b(p_e)$$

2. Bound $H(X|E, \mathcal{L})$

   By defintion of conditional entropy

   $$H(X|E, \mathcal{L}) = Pr(E = 0)H(X|E = 0, \mathcal{L}) + Pr(E = 1)H(X|E = 1, \mathcal{L})$$

   $H(X|E = 0, \mathcal{L})$ is the entropy of X given $\mathcal{L}$ when the event $X \in \mathcal{L}$ occurred which is at at most the entropy of uniform distribution on L values. Hence, $H(X|E = 0, \mathcal{L}) \leq \log L$

   $H(X|E = 1, \mathcal{L})$ is the entropy of X given $\mathcal{L}$ when the event $X \notin \mathcal{L}$ occurred which is at at most the entropy of uniform distribution on $|\mathfrak{X}| - L$ values. Hence, $H(X|E = 1, \mathcal{L}) \leq \log(|\mathfrak{X}| - L)$

   Therefore,

   $$H(X|E, \mathcal{L}) \leq (1 - p_e) \log L + p_e \log(|\mathfrak{X}| - L)$$

3. Bound $H(X|\mathcal{L})$

   By chain rule of entropy

   $$H(E|\mathcal{L}) + H(X|E, \mathcal{L}) = H(X|\mathcal{L}) + H(E|X, \mathcal{L}) = H(E, X|\mathcal{L})$$

As E is completely determined by X and $\mathcal{L}$, $H(E|X, \mathcal{L}) = 0$, so

$$\begin{aligned} H(X|\mathcal{L}) &= H(E|\mathcal{L}) + H(X|E, \mathcal{L}) \\ &\leq H_b(p_e) + (1 - p_e)\log L + p_e \log(|\mathfrak{X}| - L) \end{aligned}$$

Now, as $H_b(p_e) \leq 1$ and $(1 - p_e)\log L \leq \log L$, we have

$$p_e \geq \frac{H(X|\mathcal{L}) - 1 - \log L}{\log(|\mathfrak{X}| - L)}$$

This recovers Fano inequality when set $L = 1$

□