12. SCORE-BASED STRUCTURE LEARNING

RENATO LUI GEH
NUSP: 8536030

Abstract. This document contains the solutions to the proposed exercises from Lecture 12.

1. Solutions

Exercise 1. *Prove that the following statements are true.*

(i) $0 \leq \mathbb{H}_p(\mathcal{X}) \leq \ln |dom(\mathcal{X})|$
(ii) $\mathbb{H}_p(\mathcal{X}) = 0$ *if and only if* p *is degenerate (i.e., it assigns all mass to a single configuration).*
(iii) $\mathbb{H}_p(\mathcal{X} \cup \mathcal{Y}) = \mathbb{H}_p(\mathcal{X}) + \mathbb{H}_p(\mathcal{Y})$ *if and only if* $\mathcal{X} \perp \mathcal{Y}$ *(under* p*).*

Solution.

(i) Assume that $\mathbb{H}_p(\mathcal{X}) < 0$. Then $\sum_{\mathcal{X}} p(\mathcal{X}) \ln p(\mathcal{X}) \geq 0$. But since $0 \leq p(\mathcal{X}) \leq 1$, $\ln p(\mathcal{X}) \leq 0$. Therefore $p(\mathcal{X}) \ln p(\mathcal{X}) \leq 0$ and $\sum_{\mathcal{X}} p(\mathcal{X}) \ln p(\mathcal{X}) \leq 0$, which contradicts our earlier hypothesis. Consequently we know that $\mathbb{H}_p(\mathcal{X}) > 0$. We also know that for $p(\mathcal{X}) = 1$, $\sum_X p(\mathcal{X}) \ln p(\mathcal{X}) = 0$. Thus $\mathbb{H}_p(\mathcal{X}) \geq 0$. To find the upper bound of $\mathbb{H}_p$, we must maximize the function wrt $p(\mathcal{X})$. This is true when all events are equiprobable. Let $n = |dom(\mathcal{X})|$.

$$\max \mathbb{H}_p(\mathcal{X}) = \sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = -\sum_{i=1}^{n} \frac{1}{n}(\ln 1 - \ln n) =$$
$$= -\sum_{i=1}^{n} \frac{1}{n}(-\ln n) = (-\ln n) \underbrace{\left(-\sum_{i=1}^{n} \frac{1}{n}\right)}_{1} = \ln n =$$
$$= \ln |dom(\mathcal{X})|$$

(ii) Assume a degenerate distribution where a variable X_1 from the distribution has all mass $\Pr(X_1) = 1$ and thus all $\Pr(X_2) = \ldots = \Pr(X_n) = 0$ since $\sum_X \Pr(X) = 1$. Then we have that

$$\mathbb{H}_p(\mathcal{X}) = -\sum_{i=1}^{n} \Pr(X_i) \ln \Pr(X_i)$$
$$\mathbb{H}_p(\mathcal{X}) = 1. \ln 1 + 0. \ln 0 + \cdots + 0. \ln 0 \text{ since } p \text{ is degenerate.}$$
$$\mathbb{H}_p(\mathcal{X}) = 1 \times 0 + 0 + \cdots + 0 = 0$$

Therefore if p is degenerate, then $\mathbb{H}_p(\mathcal{X}) = 0$.

Now consider an entropy function over a distribution p and $\mathbb{H}_p(\mathcal{X}) = 0$.

$$\mathbb{H}_p(\mathcal{X}) = 0 = -\sum_{i=1}^{n} \Pr(X_i) \ln \Pr(X_i)$$

The sum of all $\Pr(X_i) \ln \Pr(X_i)$ must be zero. Since each $\Pr(X_i) \ln \Pr(X_i)$ is a non-positive number, then each term must be equal to zero. This is only true if either $\Pr(X_i) = 0$ or $\Pr(X_i) = 1$. Since $\sum_{i=1}^{n} \Pr(X_i) = 1$, there can only be one $\Pr(X_i) = 1$ and the rest will be equal to zero. But this is the definition of a degenerate probability distribution. Thus, the converse is also true.

(iii) (Incomplete solution)
A set of variables $\mathcal{X}$ is independent of another set of variables $\mathcal{Y}$ if and only if $p(\mathcal{X} \cap \mathcal{Y}) = p(\mathcal{X})p(\mathcal{Y})$. Consider an entropy function $\mathbb{H}_p$. Let $n = |\mathcal{X} \cup \mathcal{Y}|$.

$$\mathbb{H}_p(\mathcal{X} \cup \mathcal{Y}) = -\sum_{\mathcal{X} \cup \mathcal{Y}} p(\mathcal{X} \cup \mathcal{Y}) \ln p(\mathcal{X} \cup \mathcal{Y})$$

But since $\mathcal{X}$ and $\mathcal{Y}$ are independent: $p(\mathcal{X} \cup \mathcal{Y}) = p(\mathcal{X}) + p(\mathcal{Y}) + p(\mathcal{X})p(\mathcal{Y})$

$$\mathbb{H}_p(\mathcal{X} \cup \mathcal{Y}) = -\sum_{\mathcal{X} \cup \mathcal{Y}} (p(\mathcal{X}) + p(\mathcal{Y}) + p(\mathcal{X})p(\mathcal{Y})) \ln(p(\mathcal{X}) + p(\mathcal{Y}) + p(\mathcal{X})p(\mathcal{Y}))$$

□

Exercise 2. *Prove that the following statements are true*

(1) $\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) = \mathbb{I}_p(\mathcal{Y}, \mathcal{X})$
(2) $\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) = \mathbb{H}_p(\mathcal{X}) + \mathbb{H}_p(\mathcal{Y}) - \mathbb{H}_p(\mathcal{X} \cup \mathcal{Y})$
(3) $\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) \geq 0$
(4) $\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) = 0$ *iff* $\mathcal{X} \perp \mathcal{Y}$ *(under* p*)*

Solution.

(i)

$$\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) = \sum_{\mathcal{X}, \mathcal{Y}} p(\mathcal{X} \cup \mathcal{Y}) \ln \frac{p(\mathcal{X} \cup \mathcal{Y})}{p(\mathcal{X})p(\mathcal{Y})}$$

$$\mathbb{I}_p(\mathcal{Y}, \mathcal{X}) = \sum_{\mathcal{Y}, \mathcal{X}} p(\mathcal{Y} \cup \mathcal{X}) \ln \frac{p(\mathcal{Y} \cup \mathcal{X})}{p(\mathcal{Y})p(\mathcal{X})}$$

Since probability functions are commutative under product, $\mathbb{I}_p(\mathcal{X}, \mathcal{Y}) = \mathbb{I}_p(\mathcal{Y}, \mathcal{X})$.

□