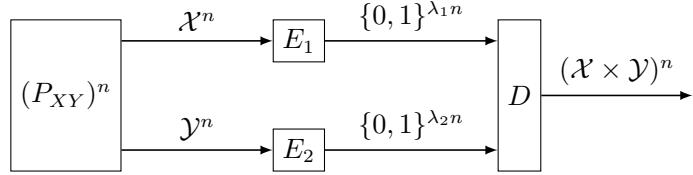


编码

请在 12 月 11 日课前提交纸质作业.

1. (10 分) 本题中, 信息熵 H 使用 2 作为底数. 令 P_{XY} 为一个支撑有限 $\mathcal{X} \times \mathcal{Y}$ 上的联合分布. 令常数 λ_1, λ_2 满足 $\lambda_1 > H(X|Y), \lambda_2 > H(Y|X), \lambda_1 + \lambda_2 > H(X, Y)$.



- (1) 请构造两个压缩函数 $E_1 : \mathcal{X}^n \rightarrow \{0, 1\}^{\lfloor \lambda_1 n \rfloor}, E_2 : \mathcal{Y}^n \rightarrow \{0, 1\}^{\lfloor \lambda_2 n \rfloor}$, 和一个解压缩函数 $D : \{0, 1\}^{\lfloor \lambda_1 n \rfloor + \lfloor \lambda_2 n \rfloor} \rightarrow (\mathcal{X} \times \mathcal{Y})^n$, 并证明

$$\Pr_{((X_1, Y_1), \dots, (X_n, Y_n)) \sim (P_{XY})^n} [D(E_1(X_1, \dots, X_n), E_2(Y_1, \dots, Y_n)) \neq ((X_1, Y_1), \dots, (X_n, Y_n))] \leq 2^{-\Theta(n)}.$$

- (2) 如果改变参数, 满足如下条件之一: a) $\lambda_1 < H(X|Y)$, b) $\lambda_2 < H(Y|X)$, c) $\lambda_1 + \lambda_2 < H(X, Y)$. 说明这时不能构造满足前一问要求的压缩函数和解压缩函数.

2. (10 分) $[\ell, n, d]$ -纠错码可以由其编码函数 $E : \{0, 1\}^n \rightarrow \{0, 1\}^\ell$ 定义, 满足

$$\forall \text{distinct } x, y \in \{0, 1\}^n, \Delta(E(x), E(y)) \geq d.$$

这里 Δ 表示汉明距离 (Hamming distance).

- (1) 证明存在常数 α . 当 $\ell > 2d$ 且 $\ell \geq 2n + \alpha d \ln(\ell/d)$ 时, 存在 $[\ell, n, d]$ -纠错码.

提示: 可以使用不等式 $\frac{\log p \cdot \log(1-p)}{\log e} \leq h(p) \leq \frac{\log p \cdot \log(1-p)}{\log 2}$. 其中 $h(p)$ 表示 $\text{Bern}(p)$ 的熵.

- (2) 证明存在常数 α . 当 $\ell > 2d$ 且 $\ell \geq n + \alpha d \ln(\ell/d)$ 时, 存在 $[\ell, n, d]$ -纠错码.

3. (5 分) $[\ell, n, d]_p$ -纠错码可以由其编码函数 $E : \mathbb{Z}_p^n \rightarrow \mathbb{Z}_p^\ell$ 定义, 满足

$$\forall \text{distinct } x, y \in \mathbb{Z}_p^n, \Delta(E(x), E(y)) \geq d.$$

这里 Δ 表示汉明距离 (Hamming distance).

- (1) (0 分) 证明当 p 为素数幂且 $p \geq \ell$ 时, 存在 $[\ell, n, \ell - n + 1]_p$ -纠错码.

- (2) 证明当 $n + d > \ell + 1$ 时, 不存在 $[\ell, n, d]_p$ -纠错码.

4. (15 分) 在有限空间 Ω 上有两个分布 P, Q . 区分器 \mathcal{D} 是一个输入域为 Ω , 输出域为 $\{0, 1\}$ 的算法 (更准确地说, 是 kernel). 我们希望让伪阳性概率 ε_{FP} 和伪阴性概率 ε_{FN} 尽量小.

$$\varepsilon_{\text{FP}} = \Pr_{X \sim P} [\mathcal{D}(X) \rightarrow 1], \quad \varepsilon_{\text{FN}} = \Pr_{X \sim Q} [\mathcal{D}(X) \rightarrow 0].$$

(1) 定义 likelihood ratio 为 $L : \Omega \rightarrow [-\infty, +\infty]$, $L(x) = \log(\frac{Q(x)}{P(x)})$.

证明: 对任何区分器 \mathcal{D} , 存在算法 $\mathcal{D}' : [-\infty, +\infty] \rightarrow \{0, 1\}$, 使得

$$\Pr_{X \sim P}[\mathcal{D}(X) \rightarrow 1] = \Pr_{X \sim P}[\mathcal{D}'(L(X)) \rightarrow 1], \quad \Pr_{X \sim Q}[\mathcal{D}(X) \rightarrow 0] = \Pr_{X \sim Q}[\mathcal{D}'(L(X)) \rightarrow 0].$$

(2) 证明: 为了最小化 $\varepsilon_{FP}, \varepsilon_{FN}$, 只须考虑如下的 likelihood ratio test 区分器 $\mathcal{D}_{\tau, \theta}$

$$\mathcal{D}_{\tau, \theta}(x) = \begin{cases} 1, & \text{if } L(x) > \tau \\ \text{Bern}(\theta), & \text{if } L(x) = \tau \\ 0, & \text{if } L(x) < \tau \end{cases}$$

(3) 改为区分 P^n 和 Q^n . 这时区分器是输入域为 Ω^n , 输出域为 $\{0, 1\}$ 的算法. 随着 n 的增长, 是否可以让 $\varepsilon_{FP}, \varepsilon_{FN}$ 分别以 $\exp(-n\alpha), \exp(-n\beta)$ 的速度趋近于 0?

具体来说, 请确定以下区域的边界

$$\left\{ (\alpha, \beta) \in \mathbb{R}_+^2 \middle| \begin{array}{l} \Pr_{X \sim P}[\mathcal{D}(X) \rightarrow 1] \leq \exp(-n\alpha) \\ \Pr_{X \sim Q}[\mathcal{D}(X) \rightarrow 0] \leq \exp(-n\beta) \end{array} \right\}$$

为了统一记号, 对任意 $\lambda \in [0, 1]$, 定义分布 P_λ 为 $P_\lambda(x) \propto (P(x))^{1-\lambda} (Q(x))^\lambda$.

提示: 上次作业第 2 题.