Bài tập về nhà số 1
Nền tảng toán học của các mô hình tạo sinh – PiMA

Chủ đề: Ôn tập xác suất thống kê

1 Phân kỳ Kullback-Leibler và bất đẳng thức Pinsker

Để đo độ khác biệt giữa hai phân phối xác suất $P(x)$ và $Q(x)$ cùng không gian mẫu $\mathcal{X}$, ta có thể sử dụng hai đại lượng sau đây (để đơn giản, ta chỉ định nghĩa chúng trên $\mathcal{X}$ rời rạc):

- Phân kỳ Kullback-Leibler: $D_{KL}(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\ln\dfrac{P(x)}{Q(x)} = \mathbb{E}_{P(x)}\left[\ln P(x) - \ln Q(x)\right]$.
- Khoảng cách L^1 giữa P và Q: $d(P,Q) = \sum_{x\in\mathcal{X}} |P(x) - Q(x)|$.

1. Để luyện tập tính toán các đại lượng này, cho $\mathcal{X} = \{1,2,3,4\}$ và các phân phối xác suất $P(x)$ và $Q(x)$ được cho như trong bảng dưới đây.

x	1	2	3	4
$P(x)$	0.25	0.25	0	0.5
$Q(x)$	0.25	0.25	0.25	0.25

   Hãy tính $D_{KL}(P\|Q)$, $D_{KL}(Q\|P)$ và $d(P,Q)$ trong trường hợp này. Quy ước $0\log 0 = 0$.

2. Chứng minh bất đẳng thức log-sum sau đây: gọi $a_1, a_2, \dots, a_n$ và $b_1, b_2, \dots, b_n$ là $2n$ số thực dương, và đặt $a = \sum_{i=1}^{n} a_i$, $b = \sum_{i=1}^{n} b_i$. Khi đó $\sum_{i=1}^{n} a_i \ln\dfrac{a_i}{b_i} \geq a\ln\dfrac{a}{b}$.

   Từ đó, hãy xác định giá trị nhỏ nhất và giá trị lớn nhất mà $D_{KL}(P\|Q)$ có thể nhận. Nêu rõ điều kiện giữa hai phân phối P và Q để đẳng thức xảy ra.

3. Chứng minh bất đẳng thức Pinsker: $D_{KL}(P\|Q) \geq \dfrac{1}{2}d^2(P,Q)$.

   Gợi ý. Có thể chứng minh thông qua hai bước sau:

   - Gọi $\mathcal{X}_1 = \{a\colon P(a) \geq Q(a)\}$, $p = \sum_{x\in\mathcal{X}_1} P(x)$ và $q = \sum_{x\in\mathcal{X}_1} Q(x)$.
   - Viết lại hiệu $D_{KL}(P\|Q) - \dfrac{1}{2}d^2(P,Q)$ theo p và q và đánh giá.