تئوری اطلاعات و کدینگ؛ جلسه 24، 6 دی 1399

* مرور جلسه قبل
* مثال هایی از رابطه آنتروپی تفاضلی و آنتروپی گسسته
* AEP برای متغیرهای تصادفی پیوسته
* آنتروپی تفاضلی مشترک و شرطی
* آنتروپی نسبی و اطلاعات متقابل برای متغیرهای پیوسته
* خاصیت بیشینه سازی توزیع مشترک گاوسی برای آنتروپی تفاضلی

---

* یادآوری:

برای متغیرهای پیوسته با تابع چگالی احتمال f_X:

$$S = \{x \mid f_X(x) > 0\}$$

$$h(X) = -\int_S f_X(x) \log f_X(x)\, dx = -\mathbb{E}[\log f_X(x)]$$

- $X \sim \text{uni}[0, a] \longrightarrow h(X) = \log a$
- $X \sim \mathcal{N}(0, \sigma^2) \longrightarrow h(X) = \frac{1}{2}\log(2\pi e\sigma^2)$

660-6769

* رابطه آنتروپی تفاضلی و آنتروپی گسسته:

$$X^{\Delta} = x_i \,, \qquad i\Delta \leqslant X < (i+1)\Delta$$: quantized version of X

$$\Rightarrow \quad H(X^{\Delta}) + \log \Delta \xrightarrow{\Delta \to 0} h(X)$$

* چند مثال درباره رابطه بالا:

• $X \sim \mathrm{uni}[0,1] \longrightarrow h(X) = 0$

می‌خواهیم با n-bit دقت X را نمایش دهیم ⟵ $\Delta = 2^{-n}$

$$\Rightarrow \quad H(X^{\Delta}) \simeq h(X) + n = n$$

در واقع اگر می‌خواستیم X را نمایش دهیم ∞ بیت لازم داشتیم. ولی برای نمایش با دقت Δ عملا به صورت زیر می‌شود:

$0.b_1 b_2 \ldots b_n$

• $X \sim uni[0, \frac{1}{8}] \rightarrow h(X) = -3$

$$H(X^\Delta) \approx -3 + n = n-3 \text{ bit}$$

در واقع: $0.000b_1 \cdots b_{n-3}$

• $X \sim N(0,100) \rightarrow h(X) = \frac{1}{2}\log(2\pi e \sigma^2) \simeq 5.37$

به طور متوسط: $\Rightarrow H(X^\Delta) = n + 5.37 \text{ bit}$

---

* AEP برای متغیرهای تصادفی پیوسته:

قضیه: اگر یک دنباله iid از متغیرهای $X_1, \dots, X_n$ با توزیع $f_X(x)$ داشته باشیم

$$\Rightarrow f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_X(x_i)$$

[illegible]

$$-\frac{1}{n}\log f(X_1,\dots,X_n) \xrightarrow[\text{in prob.}]{\text{WLLN}} \mathbb{E}\left[-\log f_X(X)\right]$$

$$= h(X)$$

* تعریف: برای هر $\varepsilon>0$ و هر n صحیح، مجموعه تیپی $A_\varepsilon^{(n)}$ نسبت به توزیع f_X به صورت زیر تعریف می‌شود:

$$A_\varepsilon^{(n)} = \left\{ (x_1,\dots,x_n)\in S^n : \left| -\frac{1}{n}\log f(x_1,\dots,x_n) - h(X)\right| \leqslant \varepsilon \right\}$$

$$= \left\{ (x_1,\dots,x_n)\in S^n : 2^{-n(h(X)+\varepsilon)} \leqslant f(x_1,\dots,x_n) \leqslant 2^{-n(h(X)-\varepsilon)} \right\}$$

* یادآوری: تعریف حجم یک مجموعه $A\subseteq\mathbb{R}^n$:

$$\mathrm{Vol}(A) \triangleq \int_A dx_1\cdots d_n$$

* قضیه: مجموعه نوعی $A_\epsilon^{(n)}$ خواص زیر را دارد:

(فرض می‌کنیم $X_1 \dots X_n$ iid و از توزیع f_X انتخاب می‌شوند)

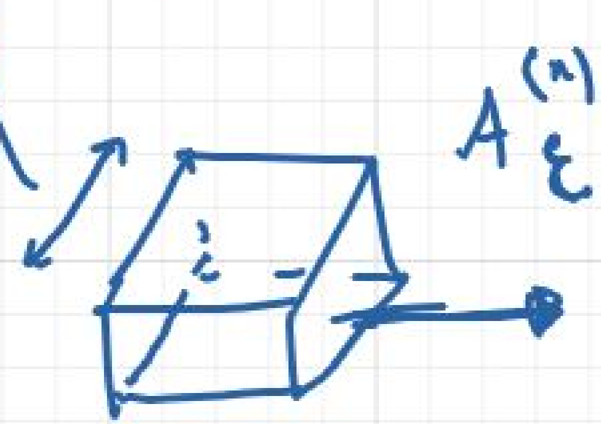
① برای n های به اندازه کافی بزرگ: $\mathbb{P}[(X_1 \cdots X_n) \in A_\epsilon^{(n)}] = \mathbb{P}[A_\epsilon^{(n)}] > 1-\epsilon$

② برای همه n ها: $Vol(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$

③ برای n های به اندازه کافی بزرگ: $Vol(A_\epsilon^{(n)}) \geq (1-\epsilon)\, 2^{n(h(X)-\epsilon)}$

$\Leftarrow$ برای n های به اندازه کافی بزرگ همه احتمال در $A_\epsilon^{(n)}$ قرار می‌گیرد و حجم $A_\epsilon^{(n)}$
ϵ های ~~به~~ به اندازه دلخواه کوچک
هم تقریبا برابر $Vol(A_\epsilon^{(n)}) \simeq 2^{nh(X)}$ است.

یعنی اینکه، مجموعه توی شبیه یک ابرمکعبی به طول ضلع $2^{h(x)}$



$A_\epsilon^{(n)}$

$vol \simeq 2^{nh(x)}$

---

* آنتروپی تفاضلی مشترک:

$f(x_1, \dots, x_n)$

$$\Rightarrow h(X_1, \dots, X_n) \triangleq -\int_S f(x_1, \dots, x_n) \log f(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

* آنتروپی شرطی:

$(X, Y) \sim f(x, y)$

$$h(X|Y) = -\int_S f(x,y) \log f(x|y)\, dx\, dy$$

(قاعده زنجیره‌ای برای چگالی احتمال

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i \mid x_1 \dots x_{i-1})$$

$$\Rightarrow \quad h(X_1, \dots, X_n) = \sum_{i=1}^{n} h(X_i \mid X_1 \dots X_{i-1})$$

* محاسبه آنتروپی تفاضلی توزیع مشترک گوسی :

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T K^{-1} (\underline{x}-\underline{\mu})\right]$$

$$\Rightarrow h(X_1, \dots, X_n) = \frac{1}{2} \log (2\pi e)^n \det(K)$$

* آنتروپی تفاضلی نسبی و اطلاعات متقابل:

$$D(f\|g) = \int_{S_f} f(x) \log \frac{f(x)}{g(x)}\, dx \ \overset{\text{Jensen's ineq}}{\geq} 0$$

متناهی است اگر که $S_f \subseteq S_g$

* تعریف اطلاعات متقابل:

$$I(X;Y) = D\big(f(x,y) \,\|\, f(x)\cdot f(y)\big) \geq 0$$

$$= \int_S f(x,y) \log\left(\frac{f(x,y)}{f(x)\,f(y)}\right) dx\,dy$$

$$= h(X) - h(X|Y)$$

$$= h(Y) - h(Y|X)$$

$$= h(X) + h(Y) - h(X,Y)$$

$$\Rightarrow \quad h(X|Y) \leq h(X)$$

$$\Rightarrow \quad h(X_1, \dots, X_n) \leq \sum_i h(X_i)$$

* ۵ تمرین (نامساوی Hadamard):

$$\det(K) \leq \prod_{i=1}^{n} K_{ii}$$

اثبات: $\underline{X} \sim N(\underline{0}, K)$

$$h(\underline{X}) = \frac{1}{2} \underbrace{\log (2\pi e)^n \det(K)}_{\frac{1}{2}\left[\log \cancel{(2\pi e)^n} + \log \det(K)\right]} \leq \sum_i h(X_i)$$

$$= \sum_i \frac{1}{2} \log (2\pi e) K_{ii}$$

$$= \frac{1}{2}\left[\log \cancel{(2\pi e)^n} + \log \prod_{i=1}^{n} K_{ii}\right]$$

$$\Rightarrow \det(K) \leq \prod_{i=1}^{n} K_{ii}$$

* خواص دیگر آنتروپی تفاضلی :

① $h(X+c) = h(X)$

عدد ثابت

② $h(a \cdot X) = h(X) + \log|a|$

عدد ثابت

②' $h(A\,\underline{X}) = h(\underline{X}) + \log|\det(A)|$

ماتریس ثابت

* قضیه: فرض کنید یک بردار تصادفی $\underline{X}$ داریم با میانگین صفر و ماتریس کوواریانس

$$K = \mathbb{E}[X X^T] \rightarrow k_{ij} = \mathbb{E}[X_i X_j]$$

آنگاه داریم:

$$h(\underline{X}) \leq \frac{1}{2} \log (2\pi e)^n \det(K)$$

شرط تساوی

$$\underline{X} \sim \mathcal{N}(0, K)$$

اثبات: فرض کنید چگالی احتمال $g(\underline{x})$ یک چگالی احتمال دلخواه با ماتریس کواریانس K باشد:

$$\int g(\underline{x})\, x_i x_j\, d\underline{x} = K_{ij}$$

$\phi_K \sim$ تابع چگالی احتمال توزیع گوسی با ماتریس کواریانس K

$$\phi_K = \frac{1}{\sqrt{(2\pi)^n \det(K)}} e^{-\frac{1}{2}\underline{x}^T K^{-1} x}$$

$$0 \leq D(g \,\|\, \phi_K)$$

$$= \int g \log (g/\phi_K)\, d\underline{x}$$

$$= \underbrace{\int g \log g \, d\underline{x}}_{-h(g)} - \int g \log \phi_K \, d\underline{x}$$

یک تابع درجه
بر حسب x_i ها

$$= \underbrace{\int g \log g \, d\underline{x}}_{-h(g)} - \int g \log \phi_K \, d\underline{x}$$

یک تابع درجه ۲
برحسب x_i ها

یک سری گشتاور دوم
توزیع g را حساب می‌کنیم.
($\mathbb{E}_g[X_iX_j]$ ها)

می‌دانیم که g و ϕ_K گشتاور دوم های یکسانی دارند. ←

$$\mathbb{E}_g[X_iX_j] = \mathbb{E}_{\phi_K}[X_iX_j] \quad \forall i,j \in [1:n]$$

$$0 \leq D(g \| \phi_K)$$

$$= -h(g) - \int \phi_K \log \phi_K \, d\underline{x}$$

$$= -h(g) + \underbrace{h(\phi_K)}_{= \frac{1}{2} \log (2\pi e)^n \det(K)} \quad \checkmark$$