

فرایندهای تصادفی : جلسه 22 و 27 اردیبهشت 1400

مطالب این جلسه:

- ادانه بدت آمار کافی
- Factorization
- آمار کافی کمین (Minimal Sufficient Statistics)

و مرور علیه قبل

$$T: \mathcal{X}^n \rightarrow \mathbb{R}^k$$

• آمار : یک تابعی از مشاهدات ما :

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

• مدل مشاهدات : $x_1, \dots, x_n$
 $x_i \sim f(x_i|\theta)$

• آمار کافی : همه اطلاعات لازم درباره θ از طریق مشاهدات $\underline{x}$ را در بر دارد.

$$\underline{x} \longleftrightarrow T(\underline{x}) \longleftrightarrow \theta$$

سزین دقیق :

Whiteboard 7912667-2884-2934 → مستقل از θ : $P[X=x | T(X)=t; \theta] = P[X=x | T(X)=t]$

$\forall x \in \mathcal{X}^n$

$\forall t: P[T(X)=t] \neq 0$

* قضیه: اگر $P(x|\theta)$ تابع چگالی احتمال یا تابع مرئی احتمال x و $Q(t|\theta)$ تابع چگالی احتمال یا مرئی احتمال $T(x)$ باشد، آنگاه

$\frac{P(x|\theta)}{Q(T(x)|\theta)}$ مستقل از θ $\Leftrightarrow T(x)$ یک آماره کافی برای θ است

برای هر $x \in \mathcal{X}^n$ باشد

* مثال: فرض $X_1, \dots, X_n \overset{\theta \in \{0,1\}}{\sim}$ متاصره و iid برتری هستند $X_i \sim \text{Bernoulli}(\theta)$

و $0 < \theta < 1$. نشان دهید $T(x) = x_1 + \dots + x_n$ یک آماره کافی برای θ است.

↓
تعداد یک‌های متاصره

$$\frac{P(\underline{x}|\theta)}{Q(T(\underline{x})|\theta)} = \frac{\prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} \quad \text{where } t = \sum_{i=1}^n x_i$$
$$t = \sum x_i \quad \sum (1-x_i) = n-t$$
$$= \frac{\theta (1-\theta)}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} = \frac{1}{\binom{n}{\sum x_i}} = \theta \text{ مستقل از } \theta$$

با استفاده از مقیاس قبلی
← $T(\underline{x})$ یک آمار کافی برای θ است.

* مثال: فرض کنید $x_1, \dots, x_n$ iid $x_i \sim \mathcal{N}(\mu, \sigma^2)$
 σ^2 : معلوم و μ : نامعلوم.

برای این مسئله میان درجه میانگینی نمونه (Sample mean)

$$T(\underline{x}) \triangleq \frac{x_1 + \dots + x_n}{n} = \bar{X}$$

فردی، تجمعی است.

$$f(\underline{x}|\mu) = \overset{=C}{\frac{1}{(\sqrt{2\pi}\sigma^2)^n}} \exp\left(-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2\right)$$

$$= C \cdot \exp\left(-\frac{\sum (x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2}\right)$$

$$\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu) = 0$$

$$= C \cdot \exp\left(-\frac{\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right)$$

$$T(\underline{x}) \sim N(\mu, \frac{\sigma^2}{n})$$

تکاملی، تجمعی، $T(\underline{x})$: $\bar{x}$

$$\frac{f(\underline{x}|\mu)}{f(T(\underline{x})|\mu)} = \frac{C \cdot \exp\left(-\frac{\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right)}{\overset{=C'}{\frac{1}{\sqrt{2\pi}\frac{\sigma^2}{n}}}} \exp\left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right) = C'' \exp\left(-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}\right)$$

مسئله از μ =

$\leftarrow T(X) = \bar{X}$ یک آمار کافی برای μ .

Factorization : * معنی

فرض کنید $f(x|\theta)$ تابع چگالی یا سری احتمال نمونه x باشد. $T(x)$ یک آمار کافی برای θ است اگر و فقط اگر $g(T(x))$ و $h(x)$ وجود داشته باشند به طوری که برای $x \in \mathcal{X}^n$ و $\theta \in \Theta$:

$$f(x|\theta) = g(T(x)|\theta) \cdot h(x)$$

* مثال: برای مدل گاوسی که در بالا بررسی شد ما داریم تابع چگالی احتمال متناهیات x را می توانیم به صورت زیر بنویسیم :

$$f(\underline{x}|\mu) = \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}\right)}_{= h(\underline{x})} \underbrace{\exp\left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right)}_{g(\underbrace{T(\underline{x})}_{=\bar{x}}|\mu)}$$

⇔ طبق قضیه factorization $T(\underline{x}) = \bar{x}$ آمار کافی برای μ است.

* مثال: فرض کنید n مشاهده iid $x_1, \dots, x_n$ ما داریم که x_i ها از یک توزیع

یکنواخت بر روی مجموعه اعداد صحیح $\{0, \dots, 9\}$ (که θ یک عدد صحیح از θ)
انتخاب شده اند.
آمار کافی برای θ کدام است؟

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & : x \in \{1, \dots, \theta\} \\ 0 & : \text{ow.} \end{cases} \quad \text{عزج } x_i :$$

$f(x|\theta) = \begin{cases} \theta^{-n} & : \forall i, x_i \in \{1, \dots, \theta\} \\ 0 & : \text{o.w.} \end{cases}$ شماره 1 تا θ است :

$N_\theta = \{1, \dots, \theta\}$, $N = \{1, 2, 3, \dots\}$: دو مجموعه متناهی و نامتناهی

$f(x|\theta) = \prod_{i=1}^n \theta^{-1} \underbrace{I_{N_\theta}(x_i)}_{\text{Indicator func.} = 1 \text{ if } x_i \in N_\theta} = \theta^{-n} \prod_{i=1}^n I_{N_\theta}(x_i)$

$$= \theta^{-n} \left(\prod_{i=1}^n I_{N_\theta}(x_i) \right) \cdot \left(\prod_{i=1}^n I_N(x_i) \right)$$

$$= I_{N_\theta}(\underbrace{T(x)}_{\rightarrow \hat{=} \max_i x_i})$$

$$= \underbrace{\theta^{-n} I_{\mathcal{N}_\theta}(T(\underline{x}))}_{g(T(\underline{x})|\theta)} \underbrace{\left(\prod_{i=1}^n I_{\mathcal{N}}(x_i; \cdot) \right)}_{h(\underline{x})}$$

طبق معنی factorization $T(\underline{x}) = \max_i x_i$ یک آمار کافی برای θ است.

* اثبات معنی factorization برای متغیرهای تصادفی گسسته:
یعنی ا دل اثبات: فرض می کنیم $T(\underline{x})$ یک آمار کافی برای θ باشد.
و نتایج زیر را استخراج می کنیم:

$$g(t|\theta) \triangleq P[T(\underline{X})=t|\theta]$$

$$h(\underline{x}) \triangleq P[\underline{X}=\underline{x} | T(\underline{X})=T(\underline{x}); \theta]:$$

به علت تعویض
آمار کافی مستقل از
 θ است.

$$= P[\underline{X}=\underline{x} | T(\underline{X})=T(\underline{x})]$$

$$\begin{aligned} f(x|\theta) &= P[X=x|\theta] \\ &= P[X=x, T(X)=T(x)|\theta] \\ &= P[T(X)=T(x)|\theta] \cdot P[X=x|T(X)=T(x); \theta] \\ &= g(T(x)|\theta) \cdot h(x) \end{aligned}$$

✓

بعثی دوم: فرضی کنیم رابطه factor شدن توزیع برقرار است وی خواهیم نتیجه بگیریم $T(X)$ یک آمار کافی برای θ است.

فرضی کنیم آمار $T(X)$ توزیع $g(t|\theta)$ را دارد. میای اینکه نشان دهیم $T(X)$ یک آمار کافی است باید ثابت کنیم $\frac{f(x|\theta)}{g(T(x)|\theta)}$ مستقل از θ است.

ابتدا فرض می‌کنیم،
 $A_{T(x)} = \{y : T(y) = T(x)\}$

داریم:
$$\frac{f(x|\theta)}{L(T(x)|\theta)} = \frac{g(T(x)|\theta) \cdot h(x)}{L(T(x)|\theta)}$$

$$= \frac{g(T(x)|\theta) \cdot h(x)}{P[T(X)=T(x)|\theta]} = \frac{g(T(x)|\theta) h(x)}{\sum_{y \in A_{T(x)}} \underbrace{P[X=y|\theta]}_{=f(y|\theta)}}$$

$$= \frac{g(T(x)|\theta) h(x)}{\sum_{y \in A_{T(x)}} \underbrace{g(T(y)|\theta) h(y)}_{=g(T(x)|\theta)}} = \frac{\cancel{g(T(x)|\theta)} h(x)}{\cancel{g(T(x)|\theta)} \sum_{y \in A_{T(x)}} h(y)}$$

$$= \frac{h(\underline{x})}{\sum_{\underline{y} \in A_T(\underline{x})} h(\underline{y})} :$$

مستقل /: Θ $\xrightarrow[\text{طیله قبل}]{\text{طبیقی مقینه}}$ $T(\underline{x})$ کی آراء
کافی کای ۱۵ است

✓