به نام زندگی

کلاس حل تمرین روز سه‌شنبه ۹/۸ ساعت 14

در بخش قبل روش تخمین بیزین را بررسی کردیم که در آن هدف، مینیمم کردن یک تابع هزینه برای تخمین پارامتر تصادفی $\underline{\theta}$ از روی مشاهدات $\underline{y}$ است.

به طوری که

$$\underline{\theta} \in \Lambda , \quad \underline{y} \in \Gamma , \quad \underline{\theta} \sim w(\underline{\theta})$$

$$\{ P_{\underline{\theta}}(\underline{y}) \; ; \; \underline{\theta} \in \Lambda \} , \quad P_{\underline{\theta}}(\underline{y}) = P(\underline{y} | \underline{\theta})$$

$\Leftarrow$ جواب تخمین $\hat{\underline{\theta}}(\underline{y})$ ای است که تابع هزینه بیزین را مینیمم می کند.

* تخمین پارامترهای غیر تصادفی (پارامترهایی که به صورت یک متغیر تصادفی با تابع چگالی احتمال مشخص قابل مدل سازی نیستند)

در این بخش می‌خواهیم پارامترهای را تخمین بزنیم که اطلاعات اولیه‌ی موجود در ارتباط با آنها محدود است. به طور خاص میزان اطلاعات اولیه‌ی ما در مورد پارامتر θ به اندازه‌ای نیست که بتوانیم تابع چگالی احتمال آن را مدل سازی کنیم. به عبارت دیگر در این مسائل تابع چگالی احتمال θ (یعنی $P(\theta)$) را نداریم

نداریم. برای تخمین $\underline{\theta}$ مشاهدات $\underline{y} \in \Gamma$ را در اختیار داریم و تابع چگالی احتمال شرطی $P(\underline{y}|\underline{\theta}) = P_{\underline{\theta}}(\underline{y})$ را می‌دانیم. به عبارت دیگر خانواده‌ای از توزیع‌های $P_{\underline{\theta}}(\underline{y})$ سروکار داریم که با پارامتر $\underline{\theta}$ اندیس شده‌اند.

می‌دانیم

$$\underline{\theta} \in \Lambda, \quad \underline{y} \in \Gamma$$

Λ را نیز می‌شناسیم.

در این بخش، تابع هزینه را تابع مربع خطای تخمین در نظر می‌گیریم.

حالت برداری

$$C(\underline{a}, \underline{\theta}) = \|\underline{a} - \underline{\theta}\|^2$$

حالت تک بعدی

$$C(a, \theta) = (a - \theta)^2$$

اما این بحث قابل تعمیم به توابع هزینه دیگر نیز هست.

براساس اصول کلی مسائل تخمین، می خواهیم روش تخمینی را پیدا کنیم که میانگین تابع هزینه را مینیمم کند.

$$\underline{\hat{\theta}}(\underline{y}) = \underset{\hat{\theta}}{\mathrm{Arg\ Min}}\ E\left\{C\left[\underline{\hat{\theta}}(\underline{y}), \underline{\theta}\right]\right\}$$

اسکالر تابعی از θ

$$= \underset{\hat{\theta}}{\mathrm{Arg\ Min}}\ E_{\theta}\left\{ \overbrace{E_{y|\theta}\left\{C\left[\underline{\hat{\theta}}(\underline{y}), \underline{\theta}\right] \,\middle|\, \underline{\theta} = \underline{\theta}\right\}}^{R_{\theta}(\hat{\theta})} \right\}$$

$$P_{\underline{\theta}}(\underline{y}) = P(\underline{y}|\underline{\theta})$$

$$= \underset{\hat{\theta}}{\mathrm{Arg\ Min}}\ E\left\{E_{\underline{\theta}}\left\{C\left[\underline{\hat{\theta}}(\underline{y}), \underline{\theta}\right] \,\middle|\, \underline{\theta}\right\}\right\}$$

برای می‌نیمم کردن میانگین تابع هزینه کافی است که تابع $R_{\underline{\theta}}(\hat{\underline{\theta}})$ را می‌نیمم کنیم.

$$R_{\underline{\theta}}(\hat{\underline{\theta}}) = E_{\underline{y}|\underline{\theta}}\left\{ C\left[\hat{\underline{\theta}}(\underline{y}), \underline{\theta}\right] \middle| \underline{\theta} = \underline{\theta} \right\}$$

در تخمین بیزین با توجه به اطلاعات اولیه که از $\underline{\theta}$ در اختیار داشتیم، توانستیم عبارت میانگین تابع هزینه را ساده کنیم و به صورت می‌نیمم تابع هزینه‌ی بیسین بنویسیم. اما در این حالت از این مرحله، جلوتر نمی‌توانیم برویم.

$\underline{y} \in \Gamma$, $\underline{\theta} \in \Lambda$, مشخص است Λ

$\{ P_{\underline{\theta}}(\underline{y}) \; ; \; \underline{\theta} \in \Lambda \}$: مجموعه خانواده توزیع های احتمال شرطی $\underline{y}$

* یک پیشنهاد برای حل این مسأله، این است که مسأله مینیمم سازی $R_\theta(\hat{\theta})$ را به ازای یک مقدار خاص از $\underline{\theta}$ مثلاً $\underline{\theta} = \underline{\theta}_0$ حل کنیم. در این صورت جواب تخمین تابعی از $\underline{\theta}_0$ خواهد بود که مطلوب ما نیست.

$$\hat{\underline{\theta}}(\underline{y}) = \operatorname{Arg}\min_{\hat{\theta}} R_\theta(\hat{\theta}) = \operatorname{Arg}\min_{\hat{\underline{\theta}}} E_{\underline{y}|\underline{\theta}} \left\{ c\left[\hat{\underline{\theta}}(\underline{y}), \underline{\theta}\right] \,\middle|\, \underline{\theta} = \underline{\theta} \right\}$$

* راه دیگر این است که یک محدودیت یا یک شرط به مسئله اضافه کنیم و مسئله

می‌نیمم سازی را در حضور این شرط یا محدودیت، حل کنیم.

شرطی که در این بخش برای حل مسئله، اضافه می‌کنیم بدون بایاس بودن تخمین

Unbiased است.

بدون بایاس بودن برای تابع تخمین $\hat{\underline{\theta}}(\underline{y})$ به صورت زیر تعریف می‌شود

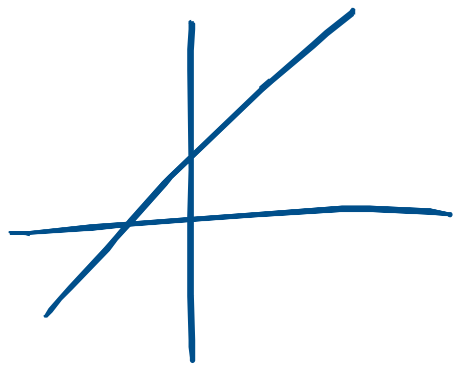
$$E_{y|\theta}\{\hat{\underline{\theta}}(\underline{y}) \mid \underline{\theta}=\underline{\theta}\} = \underline{\theta} \quad \text{یا} \quad E_{\underline{\theta}}\{\hat{\underline{\theta}}(\underline{y})\} = \underline{\theta}$$

$$E_{y|\theta}\left\{\hat{\underline{\theta}}(\underline{y}) \mid \underline{\theta} = \underline{\theta}\right\} = \int \underbrace{\hat{\underline{\theta}}(\underline{y})}\, p(\underline{y}|\underline{\theta})\, d\underline{y}$$

تابعی از $\underline{\theta}$

$$\underline{\theta} = g(\underline{\theta}) =$$ تابع از $\underline{\theta} = \underline{\theta}$

$$g(x) = ax + b$$



به این ترتیب مسأله تخمین به صورت زیر در می آید

$$\underline{\hat{\theta}}(\underline{y}) = \operatorname{Arg}\min_{\underline{\hat{\theta}}} R_{\theta}(\hat{\theta}) = \operatorname{Arg}\min_{\underline{\hat{\theta}}} E_{\underline{y}|\underline{\theta}}\left\{ C\left(\underline{\hat{\theta}}(\underline{y}), \underline{\theta}\right) \,\middle|\, \underline{\theta} = \underline{\theta} \right\}$$

$$\text{Subject to} \quad E_{\underline{\theta}}\left\{ \underline{\hat{\theta}}(\underline{y}) \right\} = \underline{\theta}$$

$$\left(E_{y|\theta}\left\{ \underline{\hat{\theta}}(\underline{y}) \,\middle|\, \underline{\theta} = \underline{\theta} \right\} = \underline{\theta} \right)$$

با در نظر گرفتن تابع هزینه مربع خطا می توان نوشت:

$$\hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\ \mathcal{E}_{\theta}\left\{ \left(\hat{\theta}(y) - \theta \right)^2 \,\middle|\, \theta = \theta \right\}$$

$$= \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\ \mathcal{E}_{\theta}\left\{ \left(\hat{\theta}(y) - \mathcal{E}_{\theta}\{\hat{\theta}(y)\} \right)^2 \,\middle|\, \theta \right\}$$

$$\mathcal{E}_{y|\theta}\left\{ \left(\hat{\theta}(y) - \mathcal{E}_{y|\theta}\{\hat{\theta}(y) \mid \theta\} \right)^2 \,\middle|\, \theta = \theta \right\}$$

$$\mathcal{E}_{\theta}\{\hat{\theta}(y)\} = \theta$$

$$\Rightarrow \hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\ \mathrm{Var}_{\theta}\left(\hat{\theta}(y) \right)$$

همان طور که می بینیم ، جوابِ تخمین ، جوابی است که واریانس خطای تخمین را می نیمم می کند. از این جهت به آن MVUE گفته می شود.

Minimum Variance Unbiased Estimation

به عبارت دیگر $\hat{\theta}(y)$ ای که $R_{\theta}(\hat{\theta})$ را می نیمم می کند، تخمینی است که واریانس خطای را می نیمم می کند.

برای به دست آوردن جواب این تخمین نیاز به بیان مفاهیمی مقدماتی داریم که در ادامه، به ترتیب این مفاهیم را معرفی خواهیم کرد.

* مفهوم کافی بودن Sufficiency یا آماره کافی

sufficient statistic

برای مشاهدات y برروی خانواده توزیع های شرطی $\{P_\theta(y), \theta \in \Lambda\}$

$T(y)$ (تابعی از y) یا آماره‌ای از y را یک آماره کافی برای تخمین θ گوییم اگر

مشاهدات $\underline{y}$ می‌گوییم اگر تمام اطلاعاتی که از $\underline{\theta}$ در $\underline{y}$ وجود دارد،

در $T(\underline{y})$ نیز وجود داشته باشد. با بیان ریاضی

$$I(\underline{\theta};\underline{y}) = I(\underline{\theta};T(\underline{y}))$$

به عبارت دیگر، $T(\underline{y})$ (آماره یا آماره‌ای از $\underline{y}$) را یک آماره کافی برای تخمین

$\underline{\theta}$ از روی مشاهدات $\underline{y}$ با خانواده توزیع‌های شرطی $\{P_{\theta}(\underline{y})\,;\,\underline{\theta}\in\Lambda\}$

می‌گوییم اگر $P_{\theta}(\underline{y}\,|\,T(\underline{y}))$ به θ بستگی نداشته باشد.

$$P_{\underline{\theta}}\left(\underline{y} \mid T(\underline{y})\right) \equiv P\left(\underline{y} \mid T(\underline{y}), \underline{\theta}\right)$$ یعنی تابع چگالی احتمال شرطی

$$= P\left(\underline{y} \mid T(\underline{y})\right)$$

به $\underline{\theta}$ بستگی نداشته باشد.

به عبارت دیگر به شرط معلوم بودن $T(\underline{y})$ ، پارامتر $\underline{\theta}$ و مشاهدات $\underline{y}$ مستقل از هم هستند. (استقلال شرطی) بنابراین اگر یکی از شرط‌های معادل استقلال شرطی نیز برای $T(\underline{y})$ برقرار باشد، می‌توان نتیجه گرفت که $T(\underline{y})$ یک آماره‌ی کافی است.

به طور مثال :

$$P(\underline{y}, \underline{\theta} \mid T(\underline{y})) = P(\underline{y} \mid T(\underline{y}))\, P(\underline{\theta} \mid T(\underline{y}))$$

* از نظر تئوری اطلاعات اگر داشته باشیم

$$I(\theta; y) = I(\theta; T(y))$$

می‌توانیم به صورت زیر رابطه بین $\underline{\theta}$ ، $\underline{y}$ و $T(\underline{y})$ را مشخص کنیم. می‌دانیم

$$\theta \rightarrow T(\underline{y}) \rightarrow \underline{y}$$

یک زنجیره مارکف است

① $$I(\theta; T(\underline{y})) \geq I(\underline{\theta}; y)$$

(به شرط داشتن $T(\underline{y})$ ، پارامتر $\underline{\theta}$ و مشاهدات $\underline{y}$ مستقل از هم هستند)

$$\underline{\theta} \rightarrow \underline{Y} \rightarrow T(\underline{Y})$$

از طرف دیگر می‌دانیم که

همواره یک زنجیره مارکف بدیهی ایجاد می‌کنند

$$I(\underline{\theta};\underline{Y}) \geq I(\underline{\theta},T(\underline{Y})) \quad ②$$

$$①, ② \Rightarrow I(\underline{\theta};T(\underline{Y})) = I(\underline{\theta};\underline{Y})$$

مثال : فرض کنیم که بردار تصادفی $\underline{X} = (X_1, X_2, \dots, X_n)^T$ یک

بردار تصادفی از متغیرهای تصادفی برنولی مستقل از هم X_i است و داریم

$X_i \sim \text{Bernolli}(\theta)$

$$X_i = \begin{cases} 1 & \text{با احتمال } \theta \\ 0 & \text{با احتمال } 1-\theta \end{cases}$$

$0 < \theta \leqslant 1$

نشان دهید که تعداد '۱' ها در بردار تصادفی $\underline{X}$ یک آماره کافی برای تخمین

θ از روی بردار مشاهدات $\underline{X}$ است.

$T(\underline{X}) =$ تعداد ۱ ها در بردار تصادفی $\underline{X}$

$$P_\theta(\underline{X} \mid T(\underline{X})) = P(\underline{X} \mid T(\underline{X}), \theta)$$ باید نشان دهیم تابع

$= P(\underline{X} \mid T(\underline{X}))$. چگالی احتمال شرطی متقابل به θ بستگی ندارد

برای اینکه مسئله را راحت تر حل کنیم، می توانیم $T(\underline{X})$ را به فرم زیر بازنویسی کنیم،

$$T(\underline{X}) = \sum_{i=1}^{n} X_i \quad , \quad X_i = \begin{cases} 1 & \theta \\ 0 & 1-\theta \end{cases}$$

از طرف دیگر

$$\underbrace{P_\theta\left(\underline{X} \,\middle|\, T(\underline{X})\right)}_{P_r\left\{\underline{X}=\underline{x} \,\middle|\, T(\underline{X})=t,\ \theta=\theta\right\}} \equiv P\left(\underline{X} \,\middle|\, T(\underline{X}),\ \theta\right)$$

$$= \frac{P_r\left\{\underline{X}=\underline{x},\ T(\underline{X})=t \,\middle|\, \theta=\theta\right\}}{P_r\left\{T(\underline{X})=t \,\middle|\, \theta=\theta\right\}} = \begin{cases} \dfrac{\cancel{\theta^{t}}\,\cancel{(1-\theta)^{n-t}}}{\binom{n}{t}\cancel{\theta^{t}}\,\cancel{(1-\theta)^{n-t}}} & T(\underline{x})=t \\[2ex] 0 & T(\underline{x})\neq t \end{cases}$$

$$P_\theta(\underline{X} \mid T(\underline{X})) \equiv P(\underline{X} \mid T(\underline{X}), \theta)$$

(به θ بستگی ندارد)

$$= \begin{cases} \dfrac{1}{\binom{n}{t}} & T(X) = t \\ 0 & \text{oth.} \end{cases}$$

$$= P(X \mid T(X))$$

با توجه به مطالب گفته شده می توان نتیجه گیری کرد که آماره کافی $T(\underline{y})$ برای تخمین $\underline{\theta}$ از روی مشاهدات $\underline{y}$ منحصر به فرد نیست. به عنوان مثال اگر $T(\underline{y})$ یک آماره کافی برای تخمین $\underline{\theta}$ روی مجموعه خانواده توزیع‌های $\{P_\theta(\underline{y}), \theta\in\Lambda\}$ باشد، هر تابع یک به یک از $T(\underline{y})$ نیز یک آماره کافی برای تخمین θ خواهد بود.

در مثال قبل دیدیم که $T(\underline{x})=\sum_{i=1}^{n} x_i$ یک آماره کافی برای تخمین θ است، بنابراین

به عنوان مثال $T^*(X) = \left(\sum_{i=1}^{n} X_i\right)^3$ نیز یک آماره کافی برای تخمین θ است.

نتیجه: هر تابع یک به یک از مشاهدات یا یک آماره کافی است و به آن آماره کافی بدیهی می‌گوییم.

در عمل ما به دنبال یک آماره کافی $T(Y)$ برای تخمین θ هستیم که فشرده‌ترین

حالت از مشاهدات را برای تخمین زدن θ ، به دست بدهد (و مطمئن هستیم

که هیچ اطلاعاتی از θ که در مشاهدات $\underline{y}$ وجود دارد، از دست نمی‌رود)

$$I(\theta; T(\underline{y})) = I(\theta; \underline{y})$$

به همین جهت آمارگان کافی به شکل تعریف می‌شود که آمارگان کافی از مشاهدات

$\underline{y}$ برای تخمین θ که فشرده‌ترین حالت مشاهدات را به دست می‌دهد.

Minimal sufficient statistic * آماره‌های کافی مینیمال

$T(\underline{Y})$ را یک آماره‌ی کافی مینیمال برای $\underline{Y}$ روی خانواده‌ی توزیع‌های مشخص

$\{P_\theta(\underline{y}), \theta \in \Omega\}$ برای تعیین می‌کنیم، اگر $T(\underline{Y})$ تابعی از هر آماره‌ی کافی

دیگر $\underline{Y}$ باشد.

با توجه به تعریف آماره‌های کامل مینیمال، پیدا کردن یک آماره‌ی کامل مینیمال

در حالت کلی برای مشاهدات y در خانواده توزیع شرطی $\{P_\theta(y), \theta \in \Lambda\}$ کار ساده‌ای

نیست. اما می‌توانیم با ایجاد بدون مفاهیم و قضیه‌هایی، آماره‌های کامل مناسب

برای تخمین θ در یک مجموعه خانواده توزیع شرطی $\{P_\theta(y), \theta \in \Lambda\}$ را به

دست بیاوریم.