به نام زندگی

* روش تخمین Baysian

همان طور که اشاره شد، در روش تخمین Baysian ، یک تابع هزینه $C[a, \theta]$ تعریف می شود که در آن منظور از $C[a, \theta]$ هزینه ای است که به ازای تخمین مقدار واقعی θ با مقدار حدس زده شده a ، می پردازیم.

بنابراین

$$C(a,\theta) : \Lambda \times \Lambda \longrightarrow \mathbb{R}^{+}$$

$$(a,\theta \in \Lambda)$$

در روش تخمین بیزین با مینیمم کردن امید ریاضی تابع هزینه و با استفاده از اطلاعات اولیه‌ای که از پارامتر θ در اختیار داریم، به یک تخمین بهینه برای پارامتر θ می‌رسیم.

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

$$a \rightarrow \hat{\theta}(y)$$

با توجه به اینکه می‌خواهیم امید ریاضی $C(\hat{\theta}(y), \theta)$ را می‌نیمم کنیم و می‌دانیم دو پارامتر تصادفی θ و y در این تابع هزینه حضور دارند.

برای می‌نیمم کردن امید ریاضی تابع هزینه، ابتدا تابع هزینه را به شکل شرطی

$R_\theta(\hat{\theta})$ را به صورت زیر تعریف می‌کنیم

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

$$= \overbrace{E_\theta\left\{ C[\hat{\theta}(y), \theta] \right\}}^{\text{نوتاسیون کتاب}} = E_\theta\left\{ C[\hat{\theta}(y), \theta] \,\middle|\, \overset{\theta=\theta}{\theta} \right\}$$

$$= \int C[\hat{\theta}(y), \theta] \underbrace{p(y|\theta)}_{p_\theta(y)} dy \equiv$$ تابعی از $\underset{\theta=\theta}{\theta}$ است

$$R_\theta(\hat{\theta}) = \underbrace{E_\theta\{C[\hat{\theta}(y),\theta]\}}_{\text{نوتاسیون کتاب}} = \underbrace{E_{y|\theta}\{C(\hat{\theta}(y),\theta)\,|\,\theta=\theta\}}_{\text{نوتاسیون آموزشی فرآیند}}$$

می دانیم که امید ریاضی تابع هزینه $C[\hat{\theta}(y),\theta]$ به صورت زیر خواهد بود

$$r(\hat{\theta}) = E\{R_\theta(\hat{\theta})\} = \overbrace{E\{E_\theta\{C[\hat{\theta}(y),\theta]\}\}}^{\text{نوتاسیون کتاب}}$$

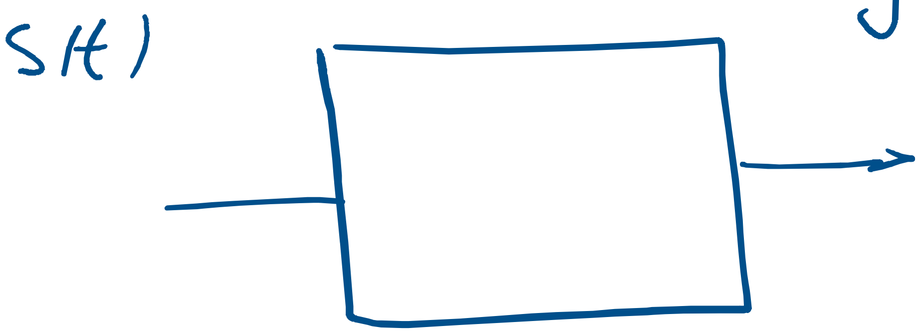
$$= E\{C(\hat{\theta}(y),\theta)\} = \underbrace{E_\theta\{E_{y|\theta}\{C(\hat{\theta}(y),\theta)\,|\,\theta=\theta\}\}}_{\text{نوتاسیون آموزشی فرآیند}}$$

در روش تخمین بیزین، باید $r(\hat{\theta})$ را می‌نیمم کنیم.

$$\hat{\theta}(y) = \operatorname{Arg}\min_{\hat{\theta}} r(\hat{\theta}) = \operatorname{Arg}\min_{\hat{\theta}} E\{C[\hat{\theta}(y), \theta]\}$$

$$\equiv \operatorname{Arg}\min_{\hat{\theta}} E\{R_{\theta}(\hat{\theta})\}$$

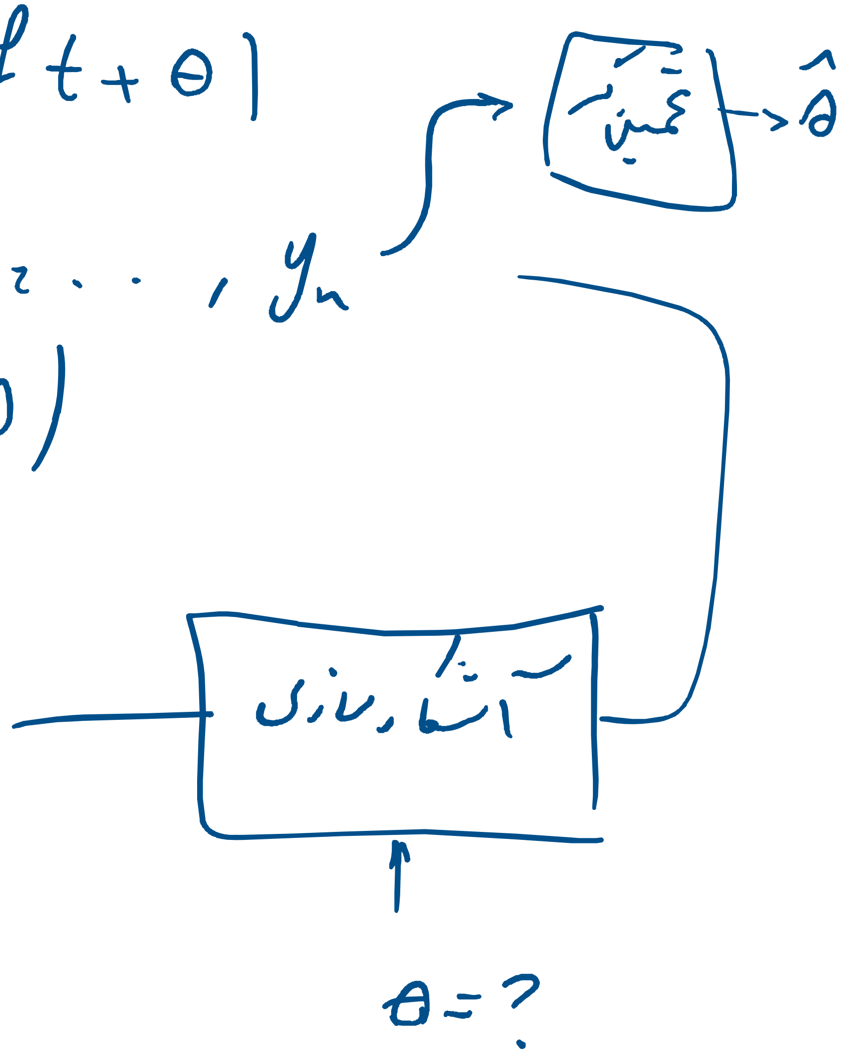
برای حل این مسأله می‌نیمم سازی، لازم است که $r(\hat{\theta})$ را با کمک قضیه احتمال کل و رابطه زنجیره‌ای بازنویسی کنیم. می‌دانیم

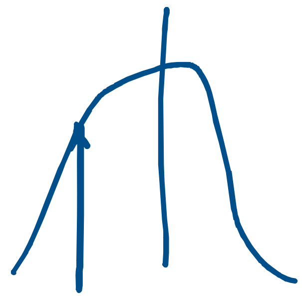


$$y = A\cos(2\pi f t + \theta)$$

$$y_1 , y_2 , \ldots , y_n$$

$$p(\underline{y} \mid \theta)$$





$$r(\hat{\theta}) = E\{R_\theta(\hat{\theta})\} = E\{E_\theta\{C[\hat{\theta}(y), \theta]\}\}$$

$$= E_\theta\{E_{y|\theta}\{C[\hat{\theta}(y), \theta] \mid \Theta = \theta\}\}$$

$$\underset{\text{رابطه زنجیره‌ای}}{=} E_y\{\underbrace{E_{\theta|y}\{C[\hat{\theta}(y), \theta] \mid y = y\}}_{\text{تابعی از } (y=y) \text{ قابل دسترسی}}\}$$

$$= E\{E_y\{C[\hat{\theta}(y), \theta]\}\} \equiv \int R_\theta(\hat{\theta})\, w(\theta)\, d\theta$$

تابع چگالی احتمال θ که در روش بیزینه به عنوان اطلاعات اولیه در دسترس است

با توجه به اینکه $C[\hat{\theta}(y),\theta]$ یک تابع غیرمنفی است، به جای مینیمم کردن

$E\{C[\hat{\theta}(y),\theta]\}$ می‌توانیم عبارت زیر را مینیمم کنیم

$$E_y\{C[\hat{\theta}(y),\theta]\} = E_{\theta|y}\{C[\hat{\theta}(y),\theta] \mid y=y\}$$

$$= \int C[\hat{\theta}(y),\theta]\, P(\theta|y)\, d\theta \equiv$$ تابعی از $y=y$ است

$$\Longrightarrow \quad \hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}} \; E_y \left\{ C[\hat{\theta}(y), \theta] \right\}$$

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

$$= \underset{\hat{\theta}}{\mathrm{Arg\,Min}} \int C[\hat{\theta}(y), \theta] \, p(\theta|y) \, d\theta$$

بنابراین در روش تخمین بیزین هدف این است که $\hat{\theta}(y)$ را به گونه‌ای پیدا کنیم که عبارت

$$E_y\{C(\hat{\theta}(y),\theta)\} = \int C(\hat{\theta}(y),\theta)\, p(\theta|y)\, d\theta$$

کمترین مقدار ممکن شود. در مسائل بیزین تابع $p_\theta(y) = p(y|\theta)$ به عنوان مدل آماری مسئله در اختیار داریم (یا از روی صورت مسئله قابل محاسبه است) تابع چگالی احتمال θ یعنی $w(\theta)$ نیز به عنوان اطلاعات اولیه از پارامتر θ در دسترس است. بنابراین اولین قدم در حل مسائل بیزین، این است که با داشتن

این اطلاعات $w(\theta|y) \equiv P(\theta|y)$ را به دست بیاوریم تا بتوانیم

در مرحله‌ی بعد مسئله‌ی بهینه سازی زیر (جواب تخمین بیزین) را به دست بیاوریم

$$\hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\ E_{y}\left\{ C\left(\hat{\theta}(y), \theta\right) \right\}$$

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

$$= \underset{\hat{\theta}}{\mathrm{Arg\,Min}} \int C\left(\hat{\theta}(y), \theta\right) w(\theta|y)\, d\theta$$

برای به دست آوردن $w(\theta|y)$ از رابطه زنجیره‌ای برای تابع چگالی احتمال شرطی کمک می‌گیریم.

رابطه زنجیره‌ای را یادآوری:

$$f_{XY}(x,y) = f_X(x)\, f_Y(y|\overset{X=x}{x}) = f_Y(y)\, f_X(x|\overset{Y=y}{y})$$

$$w(\theta|y) = \frac{w(\theta)\, p(y|\theta)}{p(y)} = \frac{\overbrace{w(\theta)}^{\text{اطلاعات اولیه}}\ \overbrace{P_\theta(y)}^{\text{مدل آماری مسئله}}}{p(y)}$$

مدل آماری مسئله (مثل کانال)

و با توجه به قضیه احتمال کلی می‌توان $P(y)$ را نیز به صورت زیر محاسبه کرد

$$P(y) = \int w(\theta)\, P_\theta(y)\, d\theta$$

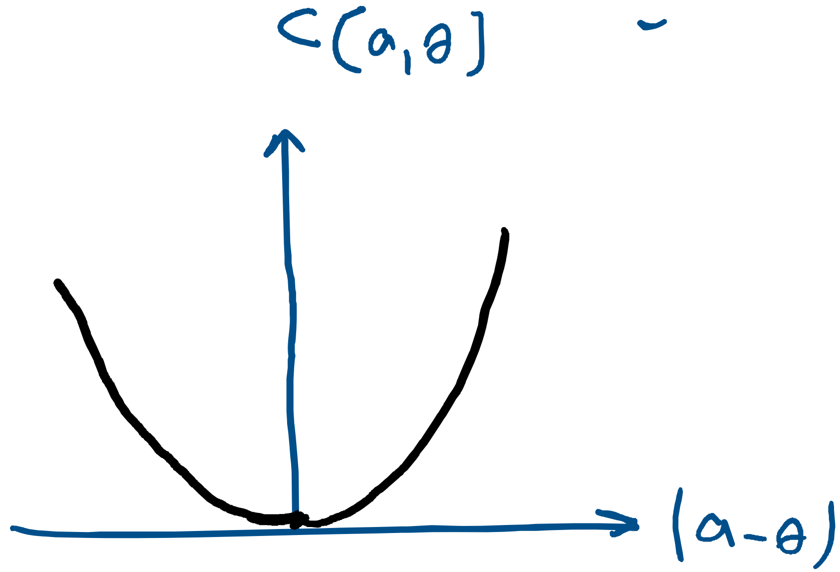
در ادامه به عنوان مثالی از روش تخمین بیزین، می‌خواهیم برای چند تابع هزینه پرکاربرد و متداول، جواب تخمین را به دست بیاوریم

* تابع هزینه مجذور خطا ← تخمین MMSE

یکی از توابع هزینه متداول و پرکاربرد در تئوری تخمین، مجذور خطای تخمین است یعنی داریم

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

$$\Lambda \times \Lambda \longrightarrow R^+ = (0, \infty)$$



$$\hat{\theta}(y) = \text{Arg}\min_{\hat{\theta}} E\left\{ C[\hat{\theta}(y), \theta] \right\}$$

$$= \text{Arg}\min_{\hat{\theta}} E\left\{ (\hat{\theta}(y) - \theta)^2 \right\}$$

همان طور که دیده می‌شود، این تخمین میانگین مربع خطا را مینیمم می‌کند. از این جهت به آن تخمین MMSE گفته می‌شود.

Minimum Mean Square Error

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

$$= \underset{\hat{\theta}}{\text{Arg Min}}\ E_y\left\{ C[\hat{\theta}(y), \theta] \right\}$$

$$= \underset{\hat{\theta}}{\text{Arg Min}}\ E_y\left\{ (\hat{\theta}(y) - \theta)^2 \right\}$$

MMSE

$$= \underset{\hat{\theta}}{\text{Arg Min}} \int_{-\infty}^{\infty} (\hat{\theta}(y) - \theta)^2\, \omega(\theta | y)\, d\theta$$

برای به دست آوردن $\hat{\theta}(y)$ باید عبارت زیر را مینیمم کنیم

$$\int_{-\infty}^{\infty} \left(\hat{\theta}(y) - \theta\right)^2 w(\theta|y)\, d\theta = E_y\left\{\left(\hat{\theta}(y) - \theta\right)^2\right\}$$

بنابراین رابطه استفاده می‌کنیم و سعی می‌کنیم با یک روش مناسب، مینیمم عبارت حاصل را به دست بیاوریم.

$$E_y\left\{\left(\hat{\theta}(y) - \theta\right)^2\right\} = E_{\theta|y}\left\{\left(\hat{\theta}(y) - \theta\right)^2 \,\middle|\, y = y\right\}$$

برای ساده کردن عبارت به این نکته توجه داریم که امید ریاضی یک عملگر خطی است.
بنابراین ابتدا از این خصوصیت امید ریاضی کمک می‌گیریم و عبارتی که بس فرامیم
می‌نسیم کنیم را ساده می‌کنیم

$$E_y\{(\hat{\theta}(y)-\theta)^2\} = E_{\theta|y}\{(\hat{\theta}(y)-\theta)^2 \mid y=y\}$$

$$= E_{\theta|y}\{(\hat{\theta}^2(y)+\theta^2-2\theta\hat{\theta}(y)) \mid y=y\}$$

$$E_y\{(\hat{\theta}(y)-\theta)^2\} = \hat{\theta}^2(y) + E_{\theta|y}\{\theta^2 \mid y=y\} - 2\hat{\theta}(y)E_{\theta|y}\{\theta \mid y=y\}$$

همان طور که می‌بینیم، $E_y\{(\hat{\theta}(y)-\theta)^2\}$ به صورت تابعی از $\hat{\theta}(y)$ قابل بیان است

$$E_y\{(\hat{\theta}(y)-\theta)^2\} = f(\hat{\theta}(y))$$

$$= \hat{\theta}^2(y) + E_{\theta|y}\{\theta^2 / y=y\} - 2\hat{\theta}(y) E\{\theta|y=y\}$$

برای اینکه عبارت $f(\hat{\theta}(y))$ را مینیمم کنیم، از آن برحسب $\hat{\theta}(y)$ مشتق می‌گیریم و حاصل را برابر صفر قرار می‌دهیم و به این ترتیب $\hat{\theta}(y)$ که $f(\hat{\theta}(y))$ را مینیمم می‌کند

به دست می آوریم. با توجه به اینکه $f(\hat{\theta}(y))$ یک تابع درجه دوم

از $\hat{\theta}(y)$ است که در آن $f''(\hat{\theta}(y)) > 0$ برقرار است، جوابی

که به دست می آوریم مینیمم (سراسری) تابع $f(\hat{\theta}(y))$ است

$$\frac{\partial f(\hat{\theta}(y))}{\partial \hat{\theta}(y)} = 2\hat{\theta}(y) - 2E_{\theta|y}\{\theta|y\} \equiv 0$$

$$\frac{\partial^2}{\partial \hat{\theta}^2(y)} f(\hat{\theta}(y)) = 2 > 0$$

$$\Longrightarrow \quad \hat{\theta}_{MMSE}(y) - E_{\theta|y}\{\theta \,|\, y = y\} = 0$$

$$\Longrightarrow \quad \hat{\theta}_{MMSE}(y) = E\{\theta \,|\, y\} = \int_{-\infty}^{\infty} \theta\, w(\theta|y)\, d\theta$$

$$= \text{Conditional Mean of } w(\theta|y)$$

= آبی از $y = y$

$\theta = 1$, $\boxed{Ey = \frac{\sum y_i}{4}}$ $(y_i = 1,\ y_2 = 1.1,\ y_3 = 0.9,\ y_4 = 1.2)$

$\hat{\theta}(y)$

$\theta = 2$ $(y_i = 2.2,\ 2.3,\ 1.8,\ 1.5)$

$\theta = 3$ $(y_i = 3.3,\ 3.5,\ 2.5,\ 2.7)$

در تخمین MMSE، حتی می‌توانیم یک فرم بسته تحلیلی را نیز به دست بیاوریم.

$$r(\hat{\theta})\Big|_{\min} \underset{\text{MMSE}}{\equiv} E\left\{\left(\hat{\theta}(y)-\theta\right)^2\right\}\Big|_{\hat{\theta}(y)=\hat{\theta}_{MMSE}}$$

$$\underset{\hat{\theta}_{MMSE}=E\{\theta|y\}}{\equiv} E\left\{\left(\theta-E\{\theta|y\}\right)^2\right\}$$

$$\Longrightarrow r(\hat{\theta})\Big|_{Min} = E\left\{(\theta - E\theta|y)^2\right\} = E\left\{\overbrace{E_y\{\theta - E\theta|y\}}^{\text{نوتاسیون کنش}}\right\}$$

$$= E_y\left\{\underbrace{E_{\theta|y}\left\{(\theta - E\{\theta|Y=y\})^2 \,\middle|\, Y=y\right\}}_{Var(\theta|y)}\right\}$$ فرآیند

تابعی از y

$$r(\hat{\theta})\Big|_{Min} = E\left\{Var(\theta|Y=y)\right\}$$