به نام زندگی

تخمین Baysian

* روش تخمین MMSE تابع هزینه $C[a,\theta] = (a-\theta)^2$

$$\hat{\theta}_{MMSE}(y) = E\{\theta | y\} \quad \text{Conditional Mean}$$

$$= \int \theta \, w(\theta|y) \, d\theta$$

$$E\{C[\hat{\theta}(y), \theta]\}\Big|_{Min} = E_y\{\overbrace{Var(\theta|y)}^{\text{تابعی از } y=y}\} = E\{Var(\theta|y)\}$$

در این جلسه می خواهیم توابع هزینه ی دیگری را در تخمین بیزین بررسی کنیم.

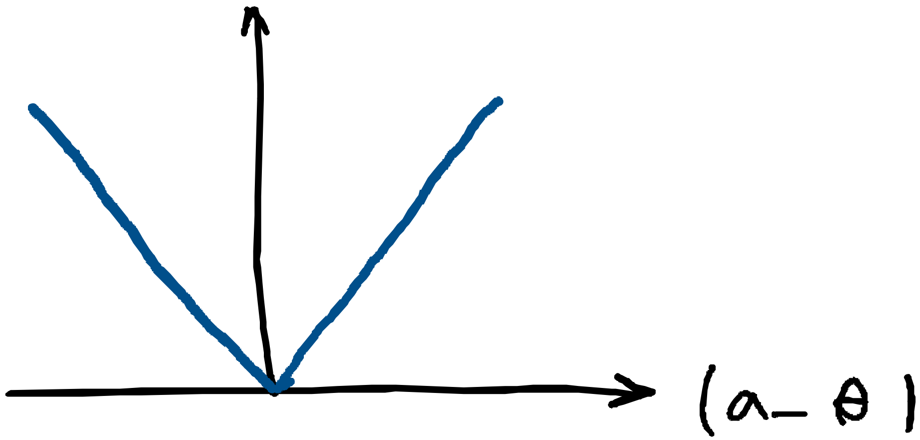
* تابع هزینه قدر مطلق خطا ← تخمین MMAE

تابع هزینه متداول دیگری که در تخمین بیزین کاربرد دارد، تابع هزینه قدر مطلق خطا است.

$$C[a,\theta] = |a-\theta|$$

$(a,\theta)\in \mathbb{R}^2 \quad , \quad C[a,\theta] : \mathbb{R}^2 \longrightarrow \mathbb{R}^+ \quad , \quad \Lambda = \mathbb{R}$

$$c(a,\theta) = |a-\theta|$$



تخمین بهترین در این حالت، تخمینی است که میانگین قدر مطلق خطا را مینیمم می‌کند

از این جهت به این تخمین

$$\hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\; E\{|\hat{\theta}(y)-\theta|\}$$

تخمین MMAE

گفته می‌شود.

Minimum Mean Absolute Error (MMAE)

$$\hat{\theta}_{MMAE}(y) = \hat{\theta}_{ABS}(y)$$

مطابق آنچه در مورد تخمین بیزین گفته شد، برای پیدا کردن تخمین، کافی است تابع هزینه بیزین (پس از مشاهده) را به صورت زیر تنظیم کنیم.

$$\hat{\theta}_{MMAE}(y) = \underset{\hat{\theta}}{\operatorname{Arg\,Min}}\ E\left\{ \left| \hat{\theta}(y) - \theta \right| \right\} = \underset{\hat{\theta}}{\operatorname{Arg\,Min}}\ E_{\theta|y}\left\{ \left| \hat{\theta}(y) - \theta \right| \,\middle|\, Y = y \right\}$$

از طرف دیگر داریم

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

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

$$= f\left(\hat{\theta}(y)\right)$$

با توجه به اینکه می خواهیم تابع هزینه پسین را مینیمم کنیم، از طرف دیگر این تابع به صورت تابعی از $\hat{\theta}(y)$ قابل بیان است، می توانیم با مشتق گیری و قرار دادن حاصل مشتق گیری برابر صفر، مقدار مینیمم را به دست بیاوریم. (باید توجه داشته باشیم که مشتق دوم مثبت باشد)

$$f(\hat{\theta}(y)) = \int_{-\infty}^{\hat{\theta}(y)} (\hat{\theta}(y)-\theta)\, w(\theta|y)\, d\theta + \int_{\hat{\theta}(y)}^{\infty} (\theta-\hat{\theta}(y))\, w(\theta|y)\, d\theta$$

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

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

یادآوری: مشتق از روابط انتگرالی

$$g(x) = \int_a^x f(t)\,dt \longrightarrow \frac{\partial}{\partial x} g(x) = f(x)$$

$$g(x) = \int_a^x f(x,t)\,dt \longrightarrow \frac{\partial}{\partial x} g(x) = f(x,x) + \int_a^x \frac{\partial}{\partial x} f(x,t)\,dt$$

$$g(x) = \int_{x}^{a} f(x,t)\,dt \longrightarrow \frac{\partial}{\partial x} g(x) = -f(x,x) + \int_{x}^{a} \frac{\partial}{\partial x} f(x,t)\,dt$$

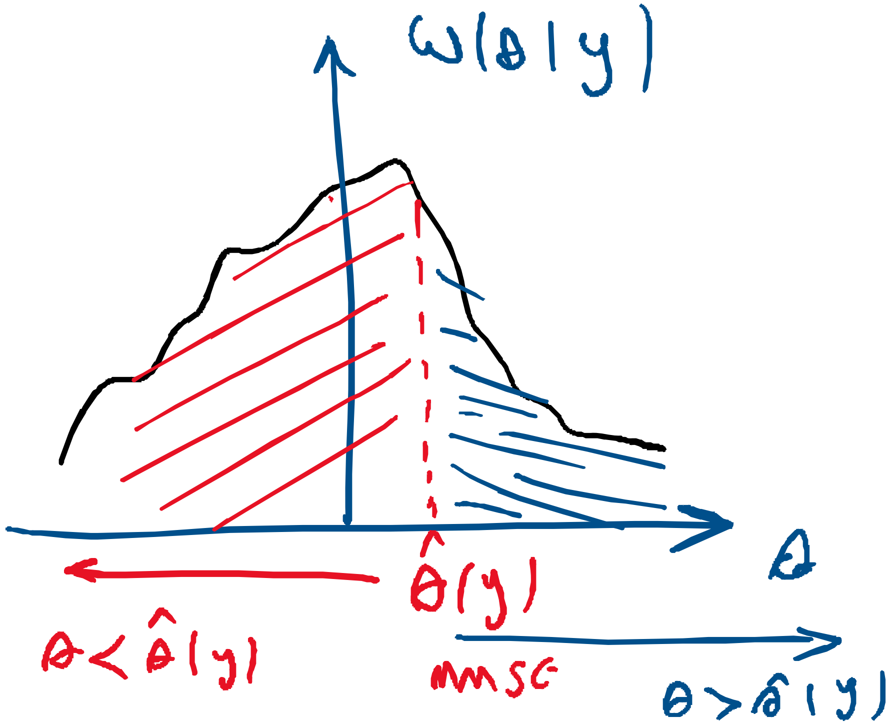
$$\Longrightarrow \frac{\partial}{\partial \hat{\theta}(y)} f(\hat{\theta}(y)) = 0 + \int_{-\infty}^{\hat{\theta}(y)} 1\, w(\theta|y)\,d\theta - 0 + \int_{\hat{\theta}(y)}^{\infty} (-1)\, w(\theta|y)\,d\theta$$

$$\frac{\partial}{\partial \hat{\theta}(y)} f(\hat{\theta}(y)) = 0 \Longrightarrow \int_{-\infty}^{\hat{\theta}(y)} w(\theta|y)\,d\theta - \int_{\hat{\theta}(y)}^{\infty} w(\theta|y)\,d\theta \Big|_{\hat{\theta}(y) = \hat{\theta}_{MMAE}} = 0$$

$$\frac{\partial^2}{\partial \hat{\theta}^2(y)} f(\hat{\theta}(y)) = w(\theta|y) + w(\theta|y)\Big|_{\hat{\theta}_{MMAE}} > 0$$

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

با توجه به اینکه مشتق دوم تابع $f(\hat{\theta}|y)$ غیرمنفی است، از حل معادله بالا به $\hat{\theta}_{MMAE}(y)$ می‌رسیم که میانگین تابع هزینه را مینیمم می‌کند.



به عبارت دیگر $\hat{\theta}_{MMAE}(y)$
میانه‌ی تابع چگالی احتمال شرطی $w(\theta|y)$
است.

$$\Longrightarrow \hat{\theta}_{MMAE}(y) = \text{Conditional Median of } w(\theta|y)$$

$$\underbrace{\int_{-\infty}^{\hat{\theta}_{MMAE}(y)} w(\theta|y)\,d\theta}_{P_r\{\theta < \hat{\theta}_{MMAE}(y)\,|\,Y=y\}} = \underbrace{\int_{\hat{\theta}_{MMAE}(y)}^{\infty} w(\theta|y)\,d\theta}_{P_r\{\theta > \hat{\theta}_{MMAE}(y)\,|\,Y=y\}}$$

• اگر تابع چگالی احتمال شرطی $w(\theta|y)$ یک تابع پیوسته از θ باشد،

پیدا کردن $\hat{\theta}_{MMAE}(y)$ که معادل میانه‌ی تابع چگالی احتمال $w(\theta|y)$ است.

به سادگی و با حل کردن یکی از معادله‌های زیر قابل محاسبه است.

$$\underbrace{\int_{-\infty}^{\hat{\theta}_{MMAE}(y)} w(\theta|y)\,d\theta}_{P_r\{\theta<\hat{\theta}(y)|Y=y\}} = \underbrace{\int_{\hat{\theta}_{MMAE}(y)}^{\infty} w(\theta|y)\,d\theta}_{P_r\{\theta>\hat{\theta}(y)|Y=y\}} = \frac{1}{2}$$

$$\hat{\theta}_{MMAE}(y) = \text{Conditional Median of } w(\theta|y)$$

درصورتی که $w(\theta|y)$ تابع چگالی احتمال پارامتر گسسته θ باشد اما $w(\theta|y)$ دارای شکلی به صورت چند ضابطه‌ای بیان شده باشد، لازم است که از تعریف کلی میانه‌ی توابع چگالی احتمال، به صورت زیر استفاده کنیم:

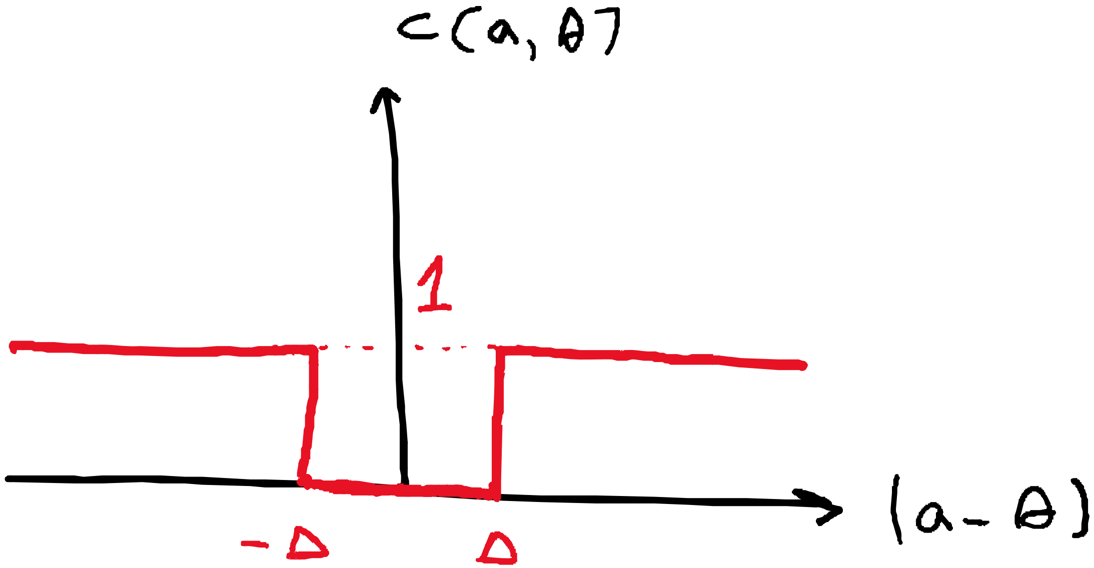
$$\begin{cases} P_r\{\theta < t \mid Y=y\} \leq P_r\{\theta > t \mid Y=y\} & t < \hat{\theta}_{MMAE}(y) \\ \\ P_r\{\theta < t \mid Y=y\} \geq P_r\{\theta > t \mid Y=y\} & t > \hat{\theta}_{MMAE}(y) \end{cases}$$

* تابع هزینه یکنواخت ← معادل جواب تخمین MAP (*)

تابع هزینه‌ی متداول دیگری که در تخمین بیزین مورد استفاده است، تابع هزینه یکنواخت، به صورت زیر است،

$$C(a,\theta) = \begin{cases} 0 & |a-\theta| \leq \Delta \\ 1 & |a-\theta| \geq \Delta \end{cases}$$

Δ معمولاً یک مقدار کوچک درنظر گرفته می‌شود.



$$\hat{\theta}(y) = \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\; E\{c(\hat{\theta}(y), \theta)\} \equiv \underset{\hat{\theta}}{\mathrm{Arg\,Min}}\; E\{c(\hat{\theta}(y), \theta) \mid Y = y\}$$

در این روش تخمین نمی‌توانیم حدس بزنیم که چه تابعی از خطا مینیمم می‌شود. ولی پس از انجام محاسبات مینیمم‌سازی به این نتیجه می‌رسیم که جواب این تخمین معادل جواب تخمین MAP است. توجه به این نکته لازم است که تخمین MAP جز تخمین بیزین نیست (به علت تفاوت در میزان اطلاعات اولیه)

برای به دست آوردن جواب تخمین، مطابق روش بیزین، تابع هزینه بیزین را می نیمم می کنیم.

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

از طرف دیگر

$$E\left\{ C[\hat{\theta}(y), \theta] \,\middle|\, y = y \right\} = \int_{-\infty}^{\infty} C[\hat{\theta}(y), \theta]\, w(\theta|y)\, d\theta$$

با توجه

$$\Rightarrow \; E\left\{ C(\hat{\theta}(y), \theta) \,\middle|\, Y=y \right\} = \underbrace{\int_{|\hat{\theta}(y)-\theta| > \Delta} w(\theta|y)\, d\theta}_{P_r\left\{ |\hat{\theta}(y)-\theta| > \Delta \,\middle|\, Y=y \right\}}$$

به نتایج جزوه

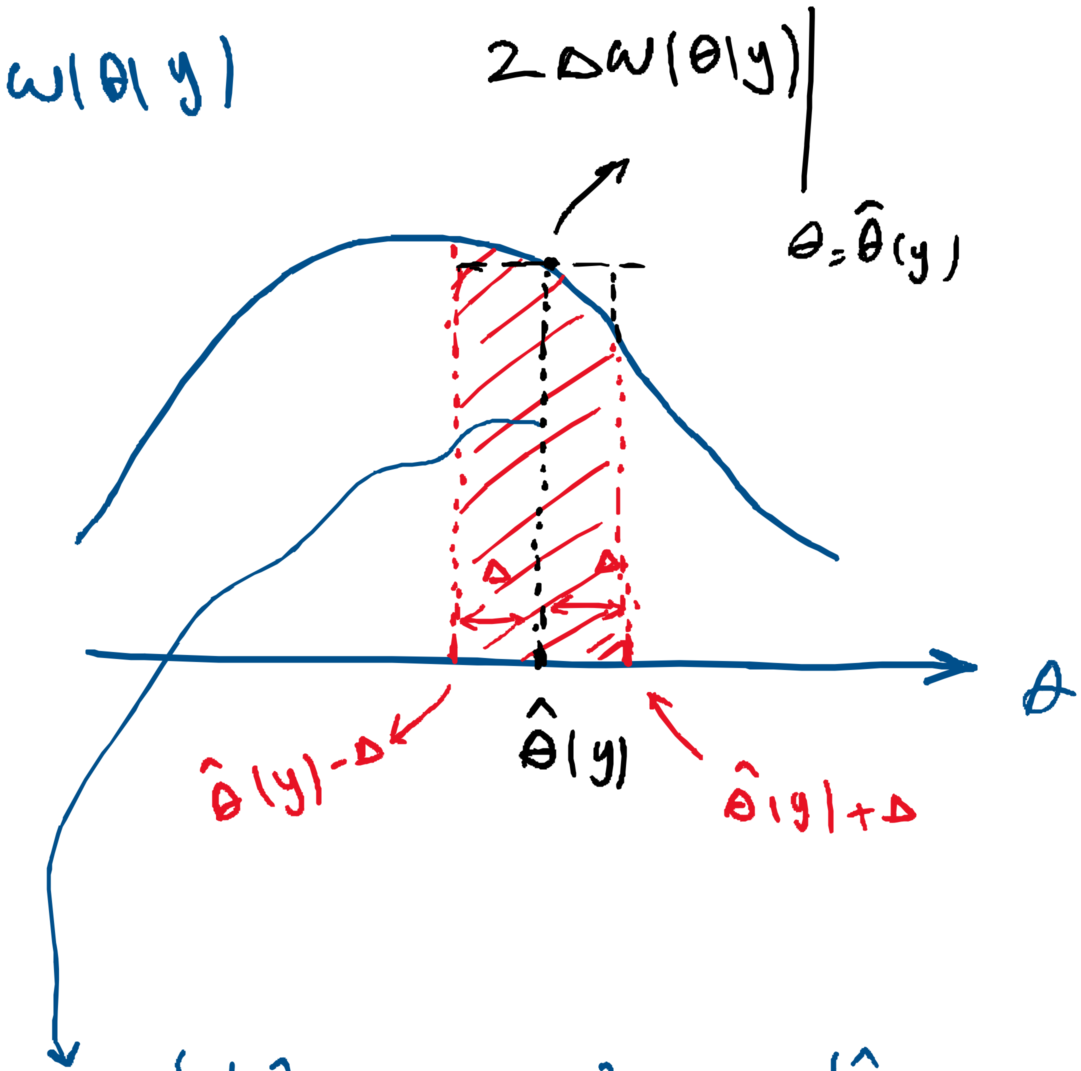
$$= 1 - P_r\left\{ |\hat{\theta}(y)-\theta| \leq \Delta \,\middle|\, Y=y \right\}$$

$$= 1 - \int_{|\hat{\theta}(y)-\theta| \leq \Delta} w(\theta|y)\, d\theta$$

$$|\hat{\theta}(y)-\theta| \leq \Delta \;\longrightarrow\; -\Delta \leq \hat{\theta}(y)-\theta \leq \Delta$$

$$\Longrightarrow E\left\{C[\hat{\theta}(y), \theta] \,\middle|\, Y=y\right\} = 1 - \underbrace{\int_{\hat{\theta}(y)-\Delta}^{\hat{\theta}(y)+\Delta} w(\theta|y)\, d\theta}_{P_r\left\{|\hat{\theta}(y)-\theta| \leq \Delta\right\}}$$

برای اینکه تابع هزینه بیسین $E\left\{C[\hat{\theta}(y), \theta] \,|\, Y=y\right\}$ را مینیمم کنیم، کافی است که $P_r\left\{|\hat{\theta}(y)-\theta| \leq \Delta\right\}$ را ماکزیمم کنیم.



با توجه به اینکه Δ یک مقدار خیلی کوچک در نظر گرفته می‌شود، برای بدست آوردن $\int_{\hat{\theta}(y)-\Delta}^{\hat{\theta}(y)+\Delta} w(\theta|y)\, d\theta$ از قضیه مقدار میانگین کمک می‌گیریم.

$$P_r\{|\hat{\theta}(y)-\theta| \leq \Delta\} = P_r\{\hat{\theta}(y)-\Delta \leq \theta \leq \hat{\theta}(y)+\Delta\}$$

$$\Rightarrow E\left\{ C\left[\hat{\theta}(y),\theta\right] \middle| Y=y \right\} = 1 - \int_{\hat{\theta}(y)-\Delta}^{\hat{\theta}(y)+\Delta} w(\theta|y)\, d\theta$$

$$= 1 - 2\Delta\, w(\theta|y)\Big|_{\theta=\hat{\theta}(y)}$$

قضیه‌ی مقدار میانگین

برای مینیمم کردن تابع هزینه‌ی بیسین، لازم است عبارت $2\Delta\, w(\theta|y)\Big|_{\theta=\hat{\theta}(y)}$ بیشترین مقدار ممکن باشد.

بنابراین $\hat{\theta}(y)$ در این حالت، نقطه‌ای است که $w(\theta|y)$ در آن نقطه ماکزیمم باشد.

$$\hat{\theta}_B(y) = \underset{\theta}{\operatorname{Arg\,Max}}\ w(\theta|y) \equiv \text{Conditional Mode of } w(\theta|y)$$

* اگر به روشهای مختلف تخمین نگاه کنیم، می‌بینیم که جواب تخمین MAP

Maximum A Posteriori probability

نیز همان Conditional Mode است. یعنی با در نظر گرفتن تابع هزینه

همانطور گفت در روش تخمین بیزین، به جوابی رسیدیم که در تخمین MAP نیز به آن می‌رسیم. با این تفاوت که میزان اطلاعات اولیه در تخمین بیزین و تخمین MAP یکسان نیست. بنابراین با وجود اینکه جواب‌ها یکسان هستند، ماهیت این دو تخمین با هم تفاوت دارد.

$$\hat{\theta}_{MAP-B}(y) = \operatorname{Arg}\max_{\theta} w(\theta|y) \equiv \text{Conditional Mode of } w(\theta|y)$$

در حالت گسسته نیز با استدلال مشابه به همین جواب خواهیم رسید.

$$\hat{\theta}_{MAP\_B}(y) = \underset{\theta}{\operatorname{Arg\,Max}}\ w(\theta|y)$$

برای پیدا کردن $w(\theta|y)$ (در حالت پیوستگی) می توانیم از تکنیک زیر استفاده کنیم.

$$w(\theta|y) = \frac{P(y|\theta)\,w(\theta)}{P(y)} = \frac{P_\theta(y)\,w(\theta)}{P(y)}$$

که در آن

$$P(y) = \int_{-\infty}^{\infty} P_\theta(y)\,w(\theta)\,d\theta$$

$$\hat{\theta}_{MAP-B}(y) = \underset{\theta}{\mathrm{Arg\,Max}}\; w(\theta|y)$$

$$= \underset{\theta}{\mathrm{Arg\,Max}}\; \frac{P_\theta(y)\, w(\theta)}{P(y)}$$

$$= \underset{\theta}{\mathrm{Arg\,Max}}\; \left\{P_\theta(y)\, w(\theta)\right\}$$

$P(y) \geq 0$ به θ بستگی ندارد

تأثیری در ماکزیمم سازی ندارد

log یک تابع اکیداً

صعودی است

$$\Longrightarrow \quad \hat{\theta}_{MAP\text{-}B}(y) = \operatorname*{Arg\,Max}_{\theta}\, w(\theta|y)$$

$$= \operatorname*{Arg\,Max}_{\theta} \left\{ \log\left(P_\theta(y)\, w(\theta)\right) \right\}$$

$$= \operatorname*{Arg\,Max}_{\theta} \left\{ \log P_\theta(y) + \log w(\theta) \right\}$$

$$\Longrightarrow \underbrace{\frac{\partial}{\partial \theta} \left\{ \log P_\theta(y) + \log w(\theta) \right\}}_{\frac{\partial^2}{\partial \theta^2} \left\{ \log P_\theta(y) + \log w(\theta) \right\} < 0} \Bigg|_{\theta = \hat{\theta}_{MAP\text{-}B}} = 0$$

$$\rightarrow \frac{\partial}{\partial \theta} \log P_\theta(y) \Big|_{\theta = \hat{\theta}_{MAP-B}} = - \frac{\partial}{\partial \theta} \log w(\theta) \Big|_{\theta = \hat{\theta}_{MAP}}$$

از حل معادله‌ی بالا می‌توانیم $\hat{\theta}_{MAP-B}(y)$ را به دست بیاوریم.

(به مشتق دوم نیز توجه داشته باشیم)

مثال : تخمین پارامتر یک تابع چگالی احتمال نمایی با روش‌های MMAE، MMSE
و MAP.B

$$P_\theta(y) = P(y|\theta) = \begin{cases} \theta e^{-\theta y} & y \geqslant 0 \\ 0 & oth. \end{cases} \qquad \theta \geqslant 0$$

$$\Gamma = \mathbb{R}, \qquad \Lambda = [0, \infty)$$

همچنین می‌دانیم که

$$w(\theta) = \begin{cases} \alpha e^{-\alpha\theta} & \theta \geqslant 0 \\ 0 & oth. \end{cases} \qquad \theta \sim \exp(\alpha)$$

* اولین قدم در حل مسائل بیزین به دست آوردن $w(\theta|y)$ است.

بنابراین ابتدا $w(\theta|y)$ را به دست می آوریم.

$$w(\theta|y) = \frac{P_\theta(y)\, w(\theta)}{P(y)} = \frac{\theta e^{-\theta y}\, \alpha e^{-\alpha\theta}}{P(y)} = \frac{\alpha\theta e^{-\theta(\alpha+y)}}{P(y)}$$

$\theta \geq 0$, $y \geq 0$

کران

$$P(y) = \int P_\theta(y)\, w(\theta)\, d\theta = \int_0^\infty \alpha\theta e^{-\theta(\alpha+y)}\, d\theta$$

یادآوری :

$$\int_0^\infty x^n e^{-\alpha x} \, dx = \frac{\Gamma(n+1)}{\alpha^{n+1}}$$

اگر $n \in \mathbb{Z}^+$ آنگاه داریم

$$\Gamma(n+1) = n!$$

$$\Rightarrow \int_0^\infty x^n e^{-\alpha x} \, dx = \frac{n!}{\alpha^{n+1}}$$

$$\Rightarrow P(y) = \alpha \int_0^\infty \theta \, e^{-(\alpha+y)\theta} \, d\theta = \alpha \frac{1!}{(\alpha+y)^2}$$

$$\Rightarrow P(y) = \frac{\alpha}{(\alpha+y)^2}$$

$$\Rightarrow w(\theta|y) = \frac{\alpha\theta \, e^{-\theta(\alpha+y)}}{P(y)} = (\alpha+y)^2 \theta e^{-\theta(\alpha+y)}$$

A) $\hat{\theta}_{MMSE}(y) = E\{\theta | y\} = \int_0^\infty \theta \, w(\theta | y) \, d\theta$

$$\Rightarrow \hat{\theta}_{MMSE}(y) = \int_0^\infty (\alpha+y)^2 \theta^2 e^{-\theta(\alpha+y)} \, d\theta$$

$$= (\alpha+y)^2 \frac{2!}{(\alpha+y)^3} = \frac{2}{(\alpha+y)} = \hat{\theta}_{MMSE}(y)$$

خطای تخمین MMSE را نیز بدست بیاورید

خطای تخمین $= E\left\{ \mathrm{Var}(\theta \mid y) \right\}$ $\qquad \mathrm{Var}\, X = E(X - m_x)^2 = EX^2 - E^2 X$

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

$$= E\left\{ E\theta^2 \mid y = y - E^2 \theta \mid y = y \right\}$$

$$= E\left\{ \int_0^{\infty} \theta^2 (\alpha + y)^2 \theta \, e^{-(\alpha+y)\theta} \, d\theta - \left(\frac{2}{\alpha + y} \right)^2 \,\middle|\, y = y \right\}$$

$$= E\left\{ (\alpha + y)^2 \frac{3!}{(\alpha + y)^4} - \left(\frac{4}{(\alpha + y)^2} \right) \,\middle|\, y = y \right\} = E\left\{ \frac{2}{(\alpha + y)^2} \,\middle|\, y = y \right\}$$

$$\Rightarrow r(\hat{\theta}_{MMSE}) = E\left\{ \frac{2}{(\alpha+y)^2} \right\}$$

$$= \int_0^\infty \frac{2}{(\alpha+y)^2}\, p(y)\, dy$$

$$= \int_0^\infty \frac{2}{(\alpha+y)^2} \frac{\alpha}{(\alpha+y)^2}\, dy = 2\alpha \left(- \frac{1}{3(\alpha+y)^3} \right)\Big|_0^\infty$$

MMSE خطای کمین $= \dfrac{2\alpha}{3\alpha^3} = \dfrac{2}{3\alpha^2} = r(\hat{\theta}_{MMSE})$