به نام خدا

تئوری تخمین (۱-۲۵۱۶۳)

تمرین سری سوم

بهار ۱۴۰۵

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

اساتید: روح الله امیری و محمّدمهدی مجاهدیان

موعد تحویل: جمعه ۸ خرداد

۱ تخمین‌گر BLUE

آ) مدلی از مشاهدات $\mathbf{x}$ را در نظر بگیرید که صرفا به ممان‌های مرتبه اول $\boldsymbol{\mu}(\boldsymbol{\theta})$ و مرتبه دوم $\mathbf{C}(\boldsymbol{\theta})$ دسترسی داریم که در آن پارامتر برداری $\boldsymbol{\theta}$ مجهول است. در کلاس تخمین‌گر BLUE را برای $\boldsymbol{\theta}$ در حالتی استخراج کردیم که $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$ باشد و $\mathbf{C}$ تابعیت $\boldsymbol{\theta}$ نداشته باشد. بیان کنید که اگر این فرضیات برقرار نباشد، چه مشکلی رخ خواهد داد؟

ب) در کلاس، تخمین‌گر BLUE را با معیار trace برای اسکالرسازی ماتریس کواریانس استخراج کردیم. در صورتی‌که معیار اسکالرسازی را به دترمینان یا λ_{max} تغییر دهیم، تخمین‌گر BLUE را برای $\boldsymbol{\theta}$ به دست بیاورید.

ج) فرض کنید $x[n] = As[n] + w[n]$ برای $n = 0, 1, \ldots, N-1$ مشاهده شود، که در آن $w[n]$ نویز با میانگین صفر و ماتریس کوواریانس $\mathbf{C}$ است و $s[n]$ یک سیگنال معلوم است. تخمین‌گر BLUE را برای پارامتر A پیدا کنید و بررسی کنید که اگر $\mathbf{s} = [s[0], s[1], \ldots s[N-1]]^\mathsf{T}$ یک بردار ویژه $\mathbf{C}$ باشد، چه اتفاقی می‌افتد. همچنین، کمترین واریانس را پیدا کنید.

د) در مساله قمت قبل، می‌توانیم بردار سیگنال $\mathbf{s}$ را به صورت ترکیب خطی از بردارهای ویژه $\mathbf{C}$ نمایش دهیم (زیرا همواره می‌توان N بردار ویژه مستقل خطی پیدا کرد.) همچنین، می‌توان نشان داد که این بردارهای ویژه به دلیل ساختار متقارن $\mathbf{C}$، متعامد هستند. بنابراین، نمایش متعامد سیگنال به صورت زیر است:

$$\mathbf{s} = \sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i$$

که در آن $\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}\}$ بردارهای ویژه متعامد $\mathbf{C}$ هستند. ثابت کنید که حداقل واریانس BLUE برابر است با:

$$\mathrm{var}(\hat{A}) = \frac{1}{\sum_{i=0}^{N-1} \frac{\alpha_i^2}{\lambda_i}}$$

که در آن λ_i مقدار ویژه $\mathbf{C}$ متناظر با بردار ویژه $\mathbf{v}_i$ است. سپس نشان دهید که انرژی سیگنال $\mathcal{E} = \mathbf{s}^T\mathbf{s}$ برابر $\sum_{i=0}^{N-1} \alpha_i^2$ است. در نهایت، ثابت کنید که اگر انرژی سیگنال محدود به مقدار $\mathcal{E}_0$ باشد، کمترین واریانس ممکن برای BLUE با انتخاب سیگنال زیر به دست می‌آید:

$$\mathbf{s} = c\mathbf{v}_{\min}$$

که در آن $\mathbf{v}_{\min}$ بردار ویژه $\mathbf{C}$ متناظر با کمترین مقدار ویژه است و c به گونه‌ای انتخاب می‌شود که شرط انرژی $\mathcal{E}_0 = \mathbf{s}^T\mathbf{s} = c^2$ برقرار شود. فرض کنید که مقادیر ویژه $\mathbf{C}$ متمایز هستند. توضیح دهید که چرا این نتیجه منطقی است.

۲ تخمین‌گر ML

آ) در کلاس دیدیم که توزیع‌های خانواده نمایی به صورت زیر بیان می‌شود:

$$f(\mathbf{x};\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\mathbf{x})\exp\left[\sum_{i=1}^{m} w_i(\boldsymbol{\theta})T_i(\mathbf{x})\right]$$

رابطه تخمین ML پارامتر $\boldsymbol{\theta}$ را استخراج کرده و آن را به صورت زیر بیان کنید:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{T}(\mathbf{x}_i) = \mathbb{E}_{\boldsymbol{\theta}}\left[\mathbf{T}(\mathbf{x})\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}}$$

که در آن $\mathbf{T}(\mathbf{x}) = [T_1(\mathbf{x}), \ldots, T_m(\mathbf{x})]^T$ بردار شامل آماره‌های کافی است. آیا برقراری رابطه فوق ملاحظاتی دارد؟ بر این اساس، چه ارتباطی بین تخمین ML و روش گشتاور وجود دارد؟

ب) فرض کنید کران کرامر-رائو توسط یک تخمین‌گر قابل حصول باشد. تحت چه شرایطی این تخمین‌گر، همان تخمین‌گر ML خواهد بود؟

ج) N مشاهده مستقل و هم‌توزیع با توزیع زیر را در نظر بگیرید:

$$f(x;\theta) = \begin{cases} \frac{x^M e^{-x/2\theta}}{(2\theta)^{M+1}M!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

که در آن M یک عدد صحیح مثبت و معلوم است. تخمین‌گر ML پارامتر θ را استخراج کرده و بایاس و واریانس آن را محاسبه کنید. همچنین، عملکرد تخمین‌گر ML را با کران کرامر-رائو (در صورت وجود) مقایسه نمایید. آیا می‌توانید تخمین‌گری بیابید که عملکرد بهتری نسبت به ML داشته باشد؟

د) قسمت قبل را برای مشاهدات مستقل و هم‌توزیع با توزیع $\mathcal{U}(0,\theta)$ تکرار نمایید.

۳ ماتریس‌های دورانی!

فرض کنید $\mathbf{R}$ یک ماتریس دورانی[1] با عناصر حقیقی به فرم زیر باشد:

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_{N-1} & r_0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_1 & \cdots & r_{N-1} & r_0 \end{bmatrix}$$

آ) نشان دهید که بردارهای ویژه تبدیل فوریه گسسته (DFT) ماتریس $\mathbf{R}$ را قطری می‌کنند:

$$\mathbf{R}\mathbf{u}_k = \lambda_k \mathbf{u}_k; \quad \sqrt{N}\mathbf{u}_k = [1,\ e^{-j2\pi k/N},\ \cdots\ e^{-j2\pi(k/N)(N-1)}]^T$$

$$\lambda_k = \sum_{n=0}^{N-1} r_n e^{-j2\pi(k/N)n}.$$

ب) اگر $\mathbf{R}$ مثبت نیمه‌معین باشد، نشان دهید $\lambda_k \geq 0$.

ج) اگر $\mathbf{R}$ متقارن باشد، نشان دهید $\lambda_k = \lambda_{N-k}$.

د) حال فرض کنید $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M)$ یک نمونه تصادفی از بردارهای تصادفی $\mathcal{N}(0, \mathbf{R})$ باشد، به طوری که $\mathbf{R}$ یک ماتریس حقیقی، متقارن، مثبت نیمه‌معین و دورانی است. کران کرامر-رائو را برای تخمین λ_k استخراج کنید.

ه) تخمین‌گر بیشینه درستنمایی (ML) λ_k را بیابید. آیا تخمین‌گر ML شما بدون بایاس، کمینه واریانس یا کارا است؟

و) تخمین‌گر ML r_n را بیابید. آیا تخمین گر ML r_n شما بدون بایاس، کیمنه واریانس یا کارا هستند؟

ز) آیا نتایج محاسبه‌شده برای $\hat{\lambda}_k$ را می‌توان به پریودوگرام مرتبط کرد؟ شهود خود را بیان کنید.

$$P_i(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1}(\mathbf{X}_i)_n e^{-jn\omega}\right|^2$$

[1] Circulant

۴ تحلیل حدی عملکرد تخمین‌گرها

در کلاس، عملکرد حدی تخمین‌گرها را با نوشتن بسط تیلور مرتبه اول تابع $\hat{\theta} = g(\mathbf{T})$ با فرض کوچک بودن واریانس آماره‌ها تحلیل کردیم. در این تمرین، قصد داریم برای تحلیل دقیق‌تر، بسط تیلور را تا مرتبه دوم در نظر بگیریم.

آ) با نوشتن بسط تیلور مرتبه دوم تابع $\hat{\theta} = g(\mathbf{T})$ حول میانگین آمارگان $\mathbb{E}\{\mathbf{T}\} = \boldsymbol{\mu}$، نشان دهید که میانگین حدی تخمین‌گر برابر است با:

$$\mathbb{E}(\hat{\theta}) = g(\boldsymbol{\mu}) + \frac{۱}{۲}\mathrm{tr}[\nabla^{۲} g(\boldsymbol{\mu})\mathbf{C}_T]$$

که در آن:

$$[\nabla^{۲} g(\boldsymbol{\mu})]_{ij} = \left.\frac{\partial^{۲} g}{\partial T_i \partial T_j}\right|_{\mathbf{T}=\boldsymbol{\mu}}$$

راهنمایی: از نتیجه زیر استفاده کنید:

$$\mathbb{E}(\mathbf{x}^{\mathsf{T}}\mathbf{y}) = \mathbb{E}(\mathrm{tr}(\mathbf{y}\mathbf{x}^{\mathsf{T}})) = \mathrm{tr}(\mathbb{E}(\mathbf{y}\mathbf{x}^{\mathsf{T}}))$$

ب) همچنین نشان دهید که واریانس تخمین‌گر به صورت حدی به صورت زیر می‌باشد:

$$\mathrm{var}(\hat{\theta}) = \nabla g(\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}_T \nabla g(\boldsymbol{\mu}) + \frac{۱}{۲}\mathrm{tr}\left[(\nabla^{۲} g(\boldsymbol{\mu})\mathbf{C}_T)^{۲}\right]$$

فرض کنید توزیع آمارگان $\mathbf{T}$ گاوسی باشد.

ج) برای مثال مشاهدات با توزیع نمایی با پارامتر مجهول λ، تخمین‌گر به روش گشتاور را بیابید و میانگین و واریانس حدی آن را بر اساس نتایج قسمت‌های قبل تحلیل کنید. این نتایج را با نتیجه تقریب مرتبه اول که در کلاس بررسی شد، مقایسه کنید. مطمئن شوید که توزیع گاوسی تقریبی $\bar{x}$ را که برای اعمال عبارت واریانس لازم است، صادق است. همچنین، نتایج را با آنچه از تئوری تخمین ML مجانبی پیش‌بینی می‌شود مقایسه کنید (توجه کنید که این تخمین‌گر، ML نیز هست).

۵ ارتباط الگوریتم EM و تکنیک MM

فرض کنید $\mathbf{X}$ دادهٔ مشاهده‌شده، $\mathbf{Z}$ متغیرهای پنهان، و $\theta \in \Theta$ پارامترها باشند. در تخمین ML هدف بیشینه‌سازی لگاریتم درست‌نمایی مشاهده‌شده است:

$$\ell(\theta) \;=\; \log p(\mathbf{X};\theta).$$

این معادل است با کمینه‌سازی $\mathcal{L}(\theta) := -\ell(\theta)$. از آنجا که

$$\mathcal{L}(\theta) \;=\; -\log \int p(\mathbf{X},\mathbf{Z};\theta)\, d\mathbf{Z},$$

انتگرال زیر لگاریتم، بهینه‌سازی مستقیم را دشوار می‌کند. روش MM[۲] یک تکنیک بهینه‌سازی است که این دشواری را با ساختن یک تابع جایگزین مافوق[۳] دور می‌زند.

طبق تعریف، تابع $u(\theta \mid \theta^{(t)})$ در نقطهٔ $\theta^{(t)}$ بر $\mathcal{L}(\theta)$ **مافوق** است اگر:

$$u(\theta \mid \theta^{(t)}) \;\geq\; \mathcal{L}(\theta) \qquad \forall\, \theta \in \Theta \tag{۱}$$

$$u(\theta^{(t)} \mid \theta^{(t)}) \;=\; \mathcal{L}(\theta^{(t)}) \tag{۲}$$

به‌روزرسانی MM عبارت است از $\theta^{(t+۱)} = \arg\min_{\theta}\, u(\theta \mid \theta^{(t)})$.

آ) با استفاده از تنها دو شرط تعریف بالا، ثابت کنید که به‌روزرسانی MM خاصیت زیر را دارد:

$$\mathcal{L}(\theta^{(t+۱)}) \;\leq\; \mathcal{L}(\theta^{(t)}).$$

یعنی تضمین می‌شود که مقدار تابع در هر گام کاهش یابد.

[۲] Majorization Minimization
[۳] Majorized Surrogate Function

ب) فرض کنید $q(\mathbf{Z})$ یک توزیع دلخواه روی متغیرهای پنهان باشد. با شروع از رابطه $\mathcal{L}(\theta) \;=\; -\log \int q(\mathbf{Z})\, \frac{p(\mathbf{X},\mathbf{Z}|\theta)}{q(\mathbf{Z})}\, d\mathbf{Z}$ و اعمال نابرابری جنسن ثابت کنید که $u(q,\theta)$ (تعریف‌شده در رابطه زیر) یک کران بالا برای $\mathcal{L}(\theta)$ است.

$$u(q,\theta) - \mathcal{L}(\theta) \;=\; D_{KL}\big(q(\mathbf{Z}) \,\|\, p(\mathbf{Z} \mid \mathbf{X},\theta)\big)$$

با استفاده از نتیجهٔ فوق و استفاده از تحدب اکید D_{KL}، نشان دهید که شرط $u(q,\theta^{(t)}) = \mathcal{L}(\theta^{(t)})$ مستلزم آن است که:

$$q^*(\mathbf{Z}) \;=\; p(\mathbf{Z} \mid \mathbf{X},\theta^{(t)}).$$

با جایگذاری $q^* = p(\mathbf{Z} \mid \mathbf{X},\theta^{(t)})$ در $u(q,\theta)$ نشان دهید که:

$$u(\theta \mid \theta^{(t)}) \;=\; -Q(\theta \mid \theta^{(t)}) + H\big[p(\mathbf{Z} \mid \mathbf{X},\theta^{(t)})\big],$$

که در آن:

$$Q(\theta \mid \theta^{(t)}) \;=\; \mathbb{E}_{\mathbf{Z}|\mathbf{X},\,\theta^{(t)}}\big[\log p(\mathbf{X},\mathbf{Z} \mid \theta)\big],$$

و $H[\cdot]$ آنتروپی است. نشان دهید که این تابع مافوق، هر دو شرط MM را برآورده می‌کند.

با نوشتن گام به‌روزرسانی MM به صورت $\theta^{(t+1)} = \arg\min_\theta\, u(\theta \mid \theta^{(t)})$ گام‌های الگوریتم EM را استخراج کرده و با رابطه ارائه‌شده در کلاس مقایسه کنید. طبیعتا، EM حالت خاصی از تکنیک MM است و همگرایی محلی آن از قسمت الف نتیجه می‌شود.

ج) توجه کنید که توابع مافوق دیگری نیز می‌توان برای $\mathcal{L}(\theta)$ تعریف کرد. به عنوان مثال، اگر $\mathcal{L}(\theta)$ دارای پیوستگی لیپشیتز با ثابت L باشد، نشان دهید که تابع جایگزین مربعی زیر هر دو شرط MM را برآورده می‌کند:

$$u_Q(\theta \mid \theta^{(t)}) \;=\; \mathcal{L}(\theta^{(t)}) + \nabla\mathcal{L}(\theta^{(t)})^\top(\theta - \theta^{(t)}) + \frac{L}{2}\|\theta - \theta^{(t)}\|^2$$

رابطه به‌روزرسانی MM متناظر را بنویسید و مشخص کنید این به‌روزرسانی معادل کدام الگوریتم بهینه‌سازی کلاسیک است.

د) در میان تمام توزیع‌های $q(\mathbf{Z})$، ثابت کنید که تابع مافوق بهینه‌شده با $q^* = p(\mathbf{Z} \mid \mathbf{X},\theta^{(t)})$، به‌صورت نقطه‌به‌نقطه کمینه است: برای هر θ و هر q دلخواه،

$$u(\theta \mid \theta^{(t)}) \;\le\; u(q,\theta) + c(q,\theta^{(t)})$$

برای ثابت مناسبی $c(q,\theta^{(t)})$ که شما باید آن را مشخص کنید. بر این اساس بیان کنید که چرا الگوریتم EM از بهترین تابع مافوق در خانوادهٔ جنسن استفاده می‌کند؟

ه) در این قسمت، قصد داریم از روش EM برای تخمین پارامترهای مدل مخلوط گاوسی روی پایگاه داده $\mathcal{D} = \{x_n\}_{n=1}^N$ استفاده کنیم. در کلاس مدل مخلوط گاوسی را با یک متغیر مخفی بررسی کردیم. در اینجا، از دو متغیر مخفی گسسته α و β برای انتخاب میانگین و واریانس مدل گاوسی استفاده می‌کنیم به‌گونه‌ای که:

$$\alpha \sim \mathrm{Cat}(p_1,\ldots,p_m), \quad \beta \sim \mathrm{Cat}(q_1,\ldots,q_l), \quad x|\alpha = i, \beta = j \sim \mathcal{N}(\mu_i, \sigma_j^2)$$

که در آن پارامترهای مدل $\boldsymbol{\theta} = [p_1,\ldots,p_m,\mu_1,\ldots,\mu_m,q_1,\ldots,q_l,\sigma_1^2,\ldots,\sigma_l^2]^T$ هستند. با استفاده از روش EM پارامترهای مدل را تخمین بزنید.

راهنمایی:

در ابتدا، توزیع مخلوط $f(x;\boldsymbol{\theta})$ را به دست آورید. سپس، فرض کنید $\boldsymbol{\theta}^{(t)}$ از تکرار قبل داده شده است. احتمال $\mathbb{P}(\alpha_n = i, \beta_n = j|x_n;\boldsymbol{\theta}^{(t)})$ را محاسبه کرده و امید لگاریتم درست‌نمایی کامل $\mathbb{E}\{\log f(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\theta})\}$ را روی توزیع $\boldsymbol{\alpha},\boldsymbol{\beta}|\mathbf{x};\boldsymbol{\theta}^{(t)}$ محاسبه کنید (گام میانگین‌گیری). نهایتا، مقدار بهینه پارامترهای مدل $\boldsymbol{\theta}$ را به گونه‌ای بیابید که عبارت محاسبه‌شده در قسمت قبل بیشینه شود (گام بیشینه‌سازی).

۶ رگرسیون خطی با داده‌های ناقص

دانشجویی قصد دارد زمان اجرای یک برنامه را تحت پارامترهای تنظیم مختلف (معادل فضای ویژگی) پیش‌بینی کند. وی برای تولید داده آموزشی، اجراهای مختلفی از این برنامه می‌گیرد ولی مجبور می‌شود اجراهای بسیار زمان‌بر را متوقف کند و در این شرایط به مقدار واقعی زمان اجرا دسترسی نخواهد داشت.

پس از تکمیل پایگاه داده $\mathcal{D} = \{(x_i,y_i)\}_{i=1}^N$، مقدار واقعی و مقدار اندازه‌گیری شده زمان اجرا برای داده iام را به ترتیب با y_i و z_i نمایش می‌دهیم. برای تعدادی از داده‌ها (اندیس‌های ۱ تا M) مقدار دقیق زمان اجرا وجود دارد یعنی $y_i = z_i$ و برای مابقی داده‌ها (اندیس‌های $M+1$ تا N) مقدار دقیق در اختیار نیست به‌گونه‌ای که $y_i = \min(z_i, c_i)$ است که c_i زمان توقف اجرا را نشان می‌دهد.

آ) قصد داریم زمان اجرا را با مدل رگرسیون خطی $z = \mu + \varepsilon$ پیش‌بینی کنیم که $\mu = \mathbf{w}^\top\mathbf{x}$ و $\varepsilon \sim \mathcal{N}(0,\sigma^2)$ می‌باشد. تابع امید لگاریتم درست‌نمایی کامل $\mathbb{E}\{\log f(\mathbf{x},\mathbf{z};\mathbf{w},\sigma^2)\}$ را محاسبه و پارامترهای بهینه مدل $(\mathbf{w},\sigma^2)$ را با بیشینه کردن این عبارت تخمین بزنید.

ب) برای مدل گاوسی ذکر شده در قسمت قبل، گزاره‌های زیر را اثبات کرده و به کمک آن پاسخ قسمت قبل را تا جای ممکن ساده نمایید.

$$\mathbb{E}[z_i|z_i \geq c_i] = \mu_i + \sigma H\left(\frac{c_i - \mu_i}{\sigma}\right)$$
$$\mathbb{E}[z_i^2|z_i \geq c_i] = \mu_i^2 + \sigma^2 + \sigma(c_i + \mu_i)H\left(\frac{c_i - \mu_i}{\sigma}\right)$$

در رابطه فوق $H(u) = \frac{\phi(u)}{1-\Phi(u)}$ که $\phi(u)$ تابع PDF و $\Phi(u)$ تابع CDF نرمال استاندارد است.

۷ تخمین حداقل مربعات ترتیبی

آ) با بلوک‌بندی مناسب ماتریس $\mathbf{H}$ (معادله مشاهدات از ابتدا تا کنون و سطر مربوط به مشاهده جدید) و استفاده از لم معکوس ماتریس، روابط ارائه‌شده در کلاس در خصوص تخمین حداقل مربعات ترتیبی را استخراج کنید. همچنین، بیان کنید این الگوریتم در کاربردهای مرتبط با فیلترهای وفقی چه استفاده‌ای دارد؟

ب) در این قسمت، مقداردهی اولیه یک تخمین‌گر حداقل مربعات ترتیبی را بررسی می‌کنیم. فرض کنید که برای مقداردهی اولیه تخمین‌گر حداقل مربعات ترتیبی، مقادیر $\hat{\boldsymbol{\theta}}[-1]$ و $\boldsymbol{\Sigma}[-1] = \alpha\mathbf{I}$ را انتخاب می‌کنیم. نشان خواهیم داد که وقتی $\alpha \to \infty$، تخمین‌گر حداقل مربعات دسته‌ای[۴] به صورت

$$\begin{aligned}\hat{\boldsymbol{\theta}}_B[n] &= \left(\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]\right)^{-1}\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{x}[n] \\ &= \left(\sum_{k=0}^{n}\frac{1}{\sigma_k^2}\mathbf{h}[k]\mathbf{h}^T[k]\right)^{-1}\left(\sum_{k=0}^{n}\frac{1}{\sigma_k^2}\mathbf{x}[k]\mathbf{h}^T[k]\right)\end{aligned}$$

برای $n \geq p$ با تخمین‌گر حداقل مربعات ترتیبی یکسان است.

ابتدا فرض کنید که بردارهای مشاهده اولیه $\{\mathbf{h}[-p], \mathbf{h}[-(p-1)], \ldots, \mathbf{h}[-1]\}$ و واریانس‌های نویز $\{\sigma_{-p}^2, \sigma_{-(p-1)}^2, \ldots, \sigma_{-1}^2\}$ وجود دارند، به طوری که برای تخمین اولیه و کوواریانس انتخاب‌شده بتوانیم از یک تخمین‌گر دسته‌ای استفاده کنیم. بنابراین

$$\hat{\boldsymbol{\theta}}[-1] = \hat{\boldsymbol{\theta}}_B[-1] = \left(\mathbf{H}^T[-1]\mathbf{C}^{-1}[-1]\mathbf{H}[-1]\right)^{-1}\mathbf{H}^T[-1]\mathbf{C}^{-1}[-1]\mathbf{x}[-1]$$

و

$$\boldsymbol{\Sigma}[-1] = \left(\mathbf{H}^T[-1]\mathbf{C}^{-1}[-1]\mathbf{H}[-1]\right)^{-1}$$

که در آن داریم:

$$\mathbf{H}[-1] = \begin{bmatrix}\mathbf{h}^T[-p] \\ \mathbf{h}^T[-(p-1)] \\ \vdots \\ \mathbf{h}^T[-1]\end{bmatrix}, \quad \mathbf{C}[-1] = \operatorname{diag}\left(\sigma_{-p}^2, \sigma_{-(p-1)}^2, \ldots, \sigma_{-1}^2\right)$$

بنابراین، می‌توانیم تخمین اولیه برای حداقل مربعات ترتیبی را نتیجه اعمال یک تخمین‌گر دسته‌ای روی بردارهای مشاهده اولیه در نظر بگیریم. از آنجا که حداقل مربعات ترتیبی که با استفاده از یک تخمین‌گر دسته‌ای مقداردهی اولیه شده است با حداقل مربعات دسته‌ای که از همه بردارهای مشاهده استفاده می‌کند یکسان است، برای حداقل مربعات ترتیبی با شرایط اولیه مفروض داریم:

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\sum_{k=-p}^{n}\frac{1}{\sigma_k^2}\mathbf{h}[k]\mathbf{h}^T[k]\right)^{-1}\left(\sum_{k=-p}^{n}\frac{1}{\sigma_k^2}\mathbf{x}[k]\mathbf{h}^T[k]\right).$$

ثابت کنید که این عبارت را می‌توان به صورت زیر بازنویسی کرد:

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\boldsymbol{\Sigma}^{-1}[-1] + \sum_{k=0}^{n}\frac{1}{\sigma_k^2}\mathbf{h}[k]\mathbf{h}^T[k]\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}[-1]\hat{\boldsymbol{\theta}}[-1] + \sum_{k=0}^{n}\frac{1}{\sigma_k^2}\mathbf{x}[k]\mathbf{h}^T[k]\right).$$

سپس بررسی کنید که وقتی $\alpha \to \infty$ برای $0 \leq n \leq p-1$ و $n \geq p$ چه اتفاقی می‌افتد؟ این روش مقداردهی اولیه را با روش ارائه‌شده در کلاس از نظر سرعت همگرایی به $\hat{\boldsymbol{\theta}}_B[n]$ مقایسه کنید.

[۴] Batch

۸ تخمین‌گر حداقل مربعات غیرخطی

سیگنالی از یک منبع در موقعیت مجهول $\mathbf{x} \in \mathbb{R}^3$ تشعشع می‌شود و توسط سنسورهای معلوم در موقعیت‌های $\mathbf{y}_1, ..., \mathbf{y}_n \in \mathbb{R}^3$ دریافت می‌شود. با توجه به قدرت سیگنال دریافتی می‌توان تخمینی نویزی به صورت d_k از فاصله هر سنسور تا منبع به دست آورد. در مساله موقعیت‌یابی، مربع خطای مقدار واقعی و مقدار اندازه‌گیری شده به صورت زیر کمینه می‌شود:

$$\mathcal{P}_1 : \min_{\mathbf{x}} f_0(\mathbf{x}) = \sum_{k=1}^{n} (\|\mathbf{x} - \mathbf{y}_k\|_2 - d_k)^2$$

این مساله، یک مساله حداقل مربعات غیرخطی است که مساله‌ای غیرمحدب بوده و حل آن با روش‌های عددی می‌تواند موجب همگرایی به نقاط بهینه محلی شود. برای رفع این مشکل، ابتدا مساله را با تعریف متغیر کمکی $t = \mathbf{x}^\mathsf{T}\mathbf{x}$ به صورت مساله مقید زیر بیان می‌کنیم و سپس با استفاده از تحدب مخفی این مساله، پاسخ بهینه آن را خواهیم یافت.

$$\begin{aligned} \mathcal{P}_2 : \min_{\mathbf{x},t} & \sum_{k=1}^{n} (t - 2\mathbf{y}_k^\mathsf{T}\mathbf{x} + \|\mathbf{y}_k\|_2^2 - d_k^2)^2 \\ \text{s.t.} \quad & \mathbf{x}^\mathsf{T}\mathbf{x} - t = 0 \end{aligned}$$

آ) با تعریف بردارها و ماتریس‌های مناسب، مساله $\mathcal{P}_2$ را به شکل زیر بیان کنید:

$$\begin{aligned} \mathcal{P}_3 : \min_{\mathbf{z}} & \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2^2 \\ \text{s.t.} \quad & \mathbf{z}^\mathsf{T}\mathbf{C}\mathbf{z} + 2\mathbf{f}^\mathsf{T}\mathbf{z} = 0 \end{aligned}$$

ب) $\mathcal{P}_3$ جزء مسائل غیرمحدبی است که برای آن‌ها تحدب مخفی برقرار است. شرایط بهینگی (KKT) را برای $\mathcal{P}_3$ نوشته و استخراج پاسخ بهینه را تا سطح پیدا کردن ریشه‌های یک چندجمله‌ای ساده کنید.

راهنمایی: می‌توانید از تغییر متغیر $\mathbf{w} = \mathbf{Q}^\mathsf{T}\mathbf{L}^\mathsf{T}\mathbf{z}$ استفاده کنید که در آن $\mathbf{Q}$ و $\mathbf{L}$ به ترتیب ماتریس‌های بردارهای ویژه و فاکتور چالسکی به فرم $\mathbf{L}^{-1}\mathbf{C}\mathbf{L}^{-\mathsf{T}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$ و $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{L}\mathbf{L}^\mathsf{T}$ هستند.

ج) با توجه به وجود دوگانی قوی برای مساله $\mathcal{P}_3$، مساله دوگان آن را (که مساله‌ای محدب است) محاسبه کرده و پاسخ $\mathcal{P}_3$ را بر حسب پاسخ مساله دوگان بیان نمایید.

۹ سوال مطالعاتی (امتیازی)

در این سوال، قصد داریم با جنبه‌های دیگری از تخمین کلاسیک آشنا شویم. لطفا یکی از مقالات زیر را مطالعه کرده و نتایج و قضایای اصلی مطرح‌شده در آن را در حداکثر یک صفحه بیان کنید.

- این مقاله، کران کرامر–رائو و تخمین‌گر ML را در جوزه اطلاعات جانبی (قید) بررسی کرده است و خاصیت بهینگی مجانبی ML را برای تخمین‌گر مقید اثبات کرده است.

T. J., Moore, B. M. Sadler, and R. J. Kozick, "Maximum-likelihood estimation, the Cramér–Rao bound, and the method of scoring with parameter constraints," IEEE Transactions on Signal Processing, 2008.

- این مقاله، کران کرامر–رائو را با محدودیت‌هایی (قید) روی بایاس تخمین‌گرها بازنویسی کرده و یک تخمین‌گر ML تغییریافته ارائه داده است که به صورت مجانبی به این کران می‌رسد.

Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," IEEE Transactions on Signal Processing, 2004.