به نام خدا

تئوری تخمین (۲۵۱۶۳-۱)

تمرین سری سوم

بهار ۱۴۰۳-۱۴۰۴

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

اساتید: روح الله امیری و محمّدمهدی مجاهدیان

موعد تحویل: دوشنبه ۱۵ اردیبهشت

۱ تخمین‌گر BLUE

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.

۱.۱

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.

۱.۲

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

برای هر کدام از توزیع‌های زیر، تخمین‌گر ML پارامتر θ را استخراج کرده و بایاس و واریانس آن را محاسبه کنید. همچنین، عملکرد تخمین‌گر ML را با کران کرامر-رائو مقایسه نمایید. آیا می‌توانید تخمین‌گری بیابید که عملکرد بهتری نسبت به ML داشته باشد؟

۱.۲ توزیع گاما

فرض کنید دنباله‌ای از مشاهدات $x_1, x_2, ..., x_n$ داریم که مستقل و هم‌توزیع هستند و هر کدام تابع چگالی احتمال زیر را دارند:

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

که در آن M یک عدد صحیح مثبت معلوم است.

۲.۲ توزیع گاوسی

فرض کنید دنباله‌ای از مشاهدات $x_1, x_2, ..., x_n$ داشته باشیم که به صورت زیر تعریف می‌شوند:

$$x_k = \theta^{1/2} N_k, \quad k = 1, ..., n$$

که در آن $N = [N_1, ..., N_n]^\mathsf{T}$ یک بردار تصادفی گاوسی با میانگین صفر و ماتریس کوواریانس Σ است.

۳.۲ توزیع برنولی

فرض کنید $x_1, x_2, ..., x_n$ مشاهدات مستقل و هم‌توزیع از یک متغیر تصادفی برنولی با تابع جرم احتمال $P\{x=1\} = 1 - P\{x=0\} = \theta$ باشند که در آن $0 < \theta < 1$ یک پارامتر غیرتصادفی است.

۳ مدل مخلوط گاوسی

قصد داریم از روش EM برای تخمین پارامترهای مدل مخلوط گاوسی روی پایگاه داده $\mathcal{D} = \{x_n\}_{n=1}^N$ استفاده کنیم. در اینجا از دو متغیر مخفی گسسته J و K برای انتخاب میانگین و واریانس مدل گاوسی استفاده می‌کنیم به‌گونه‌ای که:

$$\begin{aligned} &J \sim \mathrm{Cat}(p_1, \dots, p_m), \\ &K \sim \mathrm{Cat}(q_1, \dots, q_l), \\ &X|J=j, K=k \sim \mathcal{N}(\mu_j, \sigma_k^2) \end{aligned}$$

که در آن پارامترهای مدل $\boldsymbol{\theta} = \{p_1, \dots, p_m, \mu_1, \dots, \mu_m, q_1, \dots, q_l, \sigma_1^2, \dots, \sigma_l^2\}$ هستند.

۱.۳

توزیع مخلوط فوق $p(x;\boldsymbol{\theta})$ را به دست آورید.

۲.۳

فرض کنید $\boldsymbol{\theta}$ از قبل داده شده است. احتمال $p(J_n=j, K_n=k|x_n;\boldsymbol{\theta})$ را محاسبه کنید.

۳.۳

امید لگاریتم درست‌نمایی کامل $\mathbb{E}_Q\{\log p(\{x_n\},\{J_n\},\{K_n\};\boldsymbol{\theta})\}$ را روی توزیع $Q(\{J_n\},\{K_n\}) = p(\{J_n\},\{K_n\}|\{x_n\};\boldsymbol{\theta})$ محاسبه کنید.

۴.۳

مقدار بهینه پارامترهای مدل $\boldsymbol{\theta}$ را به گونه‌ای بیابید که عبارت محاسبه‌شده در قسمت قبل بیشینه شود.

۴ 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 iاُم 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 $y_i = \min(z_i, c_i)$ که c_i زمان توقف اجرا را نشان می‌دهد.

۱.۴

قصد داریم زمان اجرا را با مدل رگرسیون خطی $z = \mu + \varepsilon$ پیش‌بینی کنیم که $\mu = \mathbf{w}^\mathsf{T}\mathbf{x}$ و $\varepsilon \sim \mathcal{N}(۰, \sigma^۲)$ می‌باشد. تابع امید لگاریتم درست‌نمایی کامل $\mathbb{E}\{\log p(\mathbf{x}, \mathbf{z}; \mathbf{w}, \sigma^۲)\}$ را محاسبه و پارامترهای بهینه مدل $(\mathbf{w}, \sigma^۲)$ را با بیشینه کردن این عبارت تخمین بزنید. پارامترهای حاصل باید برحسب $\mathbb{E}(z_i)$ و $\mathbb{E}(z_i^۲)$ بیان شوند.

۲.۴

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.

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

در رابطه فوق $H(u) = \frac{\phi(u)}{۱-\Phi(u)}$ که $\phi(u)$ تابع PDF و $\Phi(u)$ تابع CDF است.

۵ تخمین‌گر حداقل مربعات غیرخطی

سیگنالی از یک منبع در موقعیت مجهول $\mathbf{x} \in \mathbb{R}^۳$ تشعشع می‌شود و توسط سنسورهای معلوم در موقعیت‌های $\mathbf{y}_۱, \dots, \mathbf{y}_n \in \mathbb{R}^۳$ 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:

$$
\mathcal{P}_۱: \ \min_{\mathbf{x}} f_۰(\mathbf{x}) = \sum_{k=۱}^{n} (\|\mathbf{x} - \mathbf{y}_k\|_۲ - d_k)^۲
$$

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.

$$
\begin{aligned}
\mathcal{P}_۲: \ &\min_{\mathbf{x}, t} \sum_{k=۱}^{n} (t - ۲\mathbf{y}_k^\mathsf{T}\mathbf{x} + \|\mathbf{y}_k\|_۲^۲ - d_k^۲)^۲ \\
&\text{s.t.} \quad \mathbf{x}^\mathsf{T}\mathbf{x} - t = ۰
\end{aligned}
$$

۱.۵

با تعریف بردارها و ماتریس‌های مناسب، مساله $\mathcal{P}_۲$ را به شکل زیر بیان کنید:

$$
\begin{aligned}
\mathcal{P}_۳: \ &\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_۲^۲ \\
&\text{s.t.} \quad \mathbf{z}^\mathsf{T}\mathbf{C}\mathbf{z} + ۲\mathbf{f}^\mathsf{T}\mathbf{z} = ۰
\end{aligned}
$$

۲.۵

$\mathcal{P}_۳$ جزء مسائل غیرمحدبی است که برای آن‌ها تحدب مخفی برقرار است. شرایط بهینگی (KKT) را برای $\mathcal{P}_۳$ 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 $\mathbf{L}^{-۱}\mathbf{C}\mathbf{L}^{-\mathsf{T}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\mathsf{T}$ و $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{L}\mathbf{L}^\mathsf{T}$ هستند.

۳.۵

با توجه به وجود دوگانی قوی برای مساله $\mathcal{P}_۳$، مساله دوگان آن را (که مساله‌ای محدب است) محاسبه کرده و پاسخ $\mathcal{P}_۳$ را بر حسب پاسخ مساله دوگان بیان نمایید.

۶ 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.

۱.۶

با نوشتن بسط تیلور مرتبه دوم تابع $\hat{\theta} = g(\mathbf{T})$ حول میانیگن آماره $\boldsymbol{\mu} = \mathbb{E}\{\mathbf{T}\}$، 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:

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

۲.۶

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 آماره $\mathbf{T}$ گاوسی باشد.

۳.۶

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 MLE مجانبی پیش‌بینی می‌شود مقایسه کنید (توجه کنید که این تخمین‌گر، MLE نیز هست).