باسمه تعالی

یادگیری عمیق

پاییز ۱۴۰۲

تمرین سری اول : آمار!

---

۱) یک سکه داریم که احتمال شیر آمدن آن، p، نامشخص است. برای تخمین p سکه را n بار پرتاب کرده‌ایم که در نتیجه r بار شیر آمده است. برآوردگر بیش‌ترین درست‌نمایی (MLE) برای تخمین p را به دست آورید و اریبی و واریانس آن را محاسبه کنید.

۲) توزیع بتا خانواده‌ای از توزیع‌های احتمال پیوسته بر بازه‌ی $[0,1]$ است که با دو پارامتر مثبت α,β مشخص می‌شوند و تابع چگالی آن‌ها به صورت $f(x) = cx^{\alpha-1}(1-x)^{\beta-1}$ است. اگر در مساله‌ی قبل از روی‌کرد بیزی استفاده کنیم و توزیع پیشین p را توزیع بتا با پارامترهای α,β در نظری بگیریم، توزیع پسین را محاسبه کنید. برآوردگر بیش‌ترین احتمال پسین (MAP) را به دست آورید و درباره‌ی معنی و اثر پارامترها و نیز تعداد آزمایش‌ها توضیح دهید.

۳) در مساله‌ی برآورد هم‌زمان برای $i = 1,\cdots,n$ داریم $x_i|\mu_i \sim \mathcal{N}(\mu_i,1)$ و هدف تخمین μ_iها با مشاهده‌ی x_iها است. برآوردگر بیش‌ترین درست‌نمایی برابر است با $\mathbf{x}$ و برآوردگر جیمز-استاین با فرض $n \geq 2$ به صورت $\hat{\mu}_{JS} = [1-(n-2)/||x||^2]\mathbf{x}$ تعریف می‌شود.

الف) نشان دهید برای هر برآوردگر دل‌خواه $\hat{\mu}$ داریم $\mathbb{E}[||\hat{\mu}-\mu||^2] = \mathbb{E}[||\mathbf{x}-\mu||^2] - n + 2\sum \mathrm{cov}(\hat{\mu}_i, x_i)$

ب) به کمک انتگرال‌گیری جزء به جزء نشان دهید $\mathrm{cov}(\hat{\mu}_i, x_i) = \mathbb{E}[\partial\hat{\mu}_i/\partial x_i]$

پ) نشان دهید $\mathbb{E}[||\hat{\mu}_{JS}-\mu||^2] = n - \mathbb{E}[(n-2)^2/||x||^2]$ و نتیجه بگیرید که برای $n \geq 3$ برآوردگر جیمز-استاین با محک میانگین مربع خطا همواره از برآوردگر بیش‌ترین درست‌نمایی بهتر عمل می‌کند.

۴) فرض کنید $d \geq 2$ و $\mathbf{x}$ بردار تصادفی در فضای $\mathbb{R}^d$ باشد

الف) نشان دهید اگر x_iها گاوسی با میانگین صفر و واریانس برابر باشند توزیع $\mathbf{x}$ شعاعی است (یعنی با دوران تغییر نمی‌کند)

ب) نشان دهید اگر توزیع $\mathbf{x}$ شعاعی و مولفه‌های آن از هم مستقل باشند، x_iها توزیع گاوسی با میانگین صفر و واریانس برابر دارند.

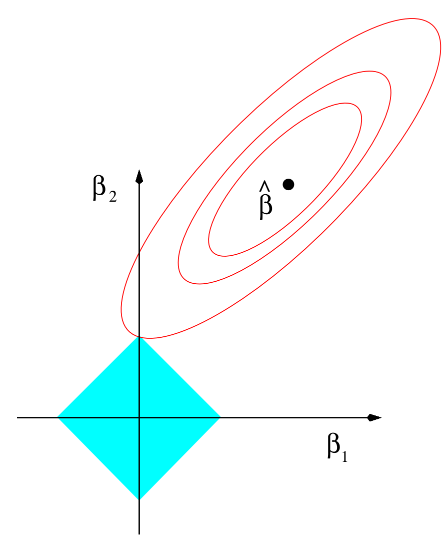
۵) در مساله‌ی رگرسیون خطی با n داده مشاهده شده و p متغیر کمکی (covariate)، فرض کنید $\mathbf{X}_{n\times p}$ ماتریس متغیرهای کمکی مشاهده شده، $\mathbf{y}_{n\times 1}$ بردار متغیرهای پاسخ مشاهده شده و $\beta_{p\times 1}$ بردار ضرایب رگرسیون باشد. طبق مدل احتمالاتی رگرسیون داریم $\mathbf{y} = \mathbf{X}\beta + \epsilon$ که $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. اگر $n > p$، ماتریس $\mathbf{X}$ رتبه‌ی کامل داشته باشد و $\hat{\beta}$ تخمین‌گر بیش‌ترین درست‌نمایی باشد و قرار دهیم $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$،

الف) نشان دهید $\hat{\mathbf{y}} - \mathbf{y} = [\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - I]\epsilon$.

ب) نشان دهید $\mathbb{E}[\|\hat{\mathbf{y}} - \mathbf{y}\|^2]/n = \sigma^2(n-p)/n$.

پ) از دو قسمت قبل درباره‌ی خطای آموزش و بیش‌برازش چه می‌فهمیم؟

۶) فرض کنید ماتریس $\mathbf{X}$ و بردار $\mathbf{y}$ با اندازه‌های $n \times p$ و $n \times 1$ داده شده‌اند. نشان دهید برای هر عدد مثبت λ عدد مثبت t وجود دارد که جواب مساله‌ی $\mathrm{argmin}_\beta \ \|\mathbf{X}\beta - \mathbf{y}\|^2 \ \text{ s.t. } \|\beta\|_p \le t$ با جواب مساله‌ی $\mathrm{argmin}_\beta \ \{\|\mathbf{X}\beta - \mathbf{y}\|^2 + \lambda\|\beta\|_p\}$ برابر است.

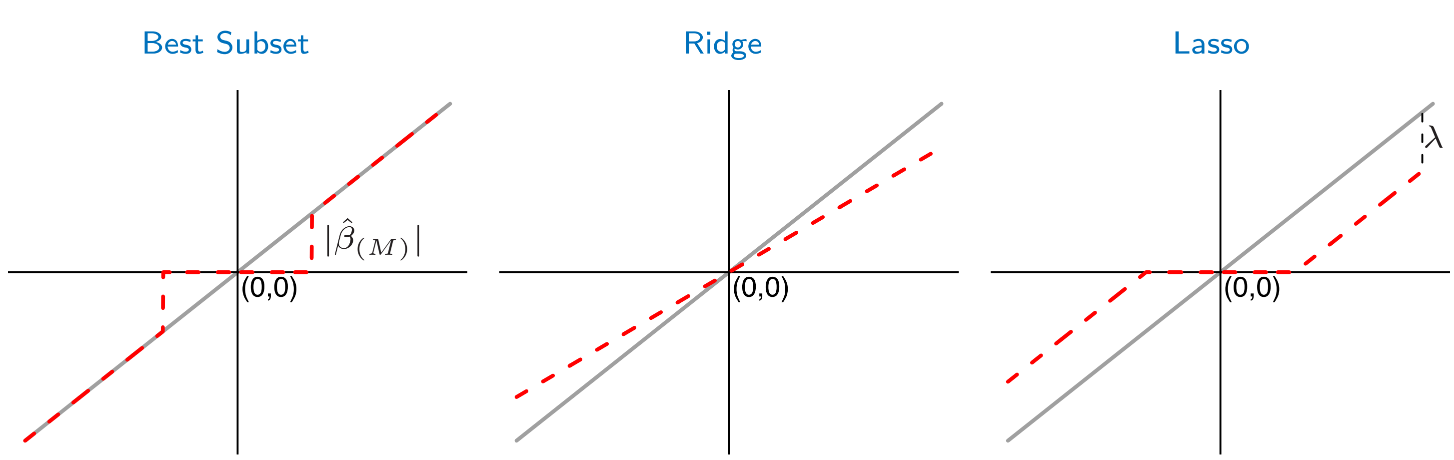


۷) فرض کنید در یک مساله‌ی رگرسیون خطی ماتریس کواریانس تجربی، $\mathbf{X}^T\mathbf{X}$، برابر با ماتریس همانی شده است! اگر $\hat{\beta}$ تخمین‌گر کم‌ترین مربعات برای مساله‌ی رگرسیون باشد نشان دهید تخمین‌ها با اضافه کردن منظم‌سازی به صورت‌های زیر تغییر می‌کنند:

الف) برای رگرسیون بهترین زیرمجموعه (Best Subset) : $\hat{\beta}_i \mathbf{1}_{\mathrm{rank}(\beta_i) \le M}$

ب) برای رگرسیون ستیغی (Ridge) : $\hat{\beta}_i/(1+\lambda)$

پ) برای رگرسیون لاسو (Lasso) : $\mathrm{sign}(\hat{\beta}_i)(|\hat{\beta}_i| - \lambda)_+$

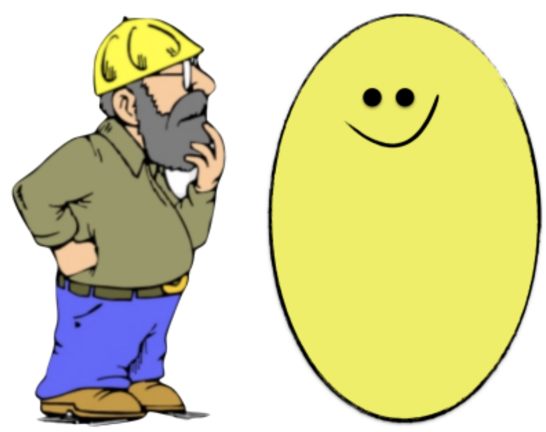


۸) از تعدادی مهندس خواسته شده که محیط یک بیضی را اندازه بگیرند و تخمین‌های $x_1, \cdots, x_n$ به دست آمده است. تخمین‌ها مستقل از هم و نااریب هستند. فرض کنید مقدار واقعی محیط برابر μ باشد و $x_i \sim \mathcal{N}(\mu, \sigma_i^2)$. چون مهندس‌ها از ابزارهای مختلفی استفاده می‌کنند σ_iها لزوما با هم برابر نیستند و مقدارشان هم نامشخص است. هدف به دست آوردن تخمین تا جای ممکن بهتری برای μ است.

الف) اگر σ_iها را بدانیم چه تخمین‌گری برای μ پیشنهاد می‌کنید؟ چرا؟

ب) تابع درست‌نمایی را بنویسید و برای به دست آوردن برآوردگر بیش‌ترین درست‌نمایی تلاش کنید.

پ) به روش بیزی عمل کنید: λ_i را معکوس σ_i^2 بگیرید. در توزیع پیشین فرض کنید μ و λ_iها مستقل‌اند و $\lambda_1, \cdots, \lambda_n \sim \Gamma(a, b)$. (تابع چگالی توزیع گاما به شکل $g(\lambda) = c\lambda^{a-1}e^{-b\lambda}$ است) هم‌چنین توزیع پیشین μ را گاوسی با میانگین صفر و واریانس بی‌نهایت (واریانسی که به بی‌نهایت میل می‌کند) بگیرید. توزیع پسین را محاسبه کنید و برآوردگر بیشترین احتمال پسین را به دست آورید.

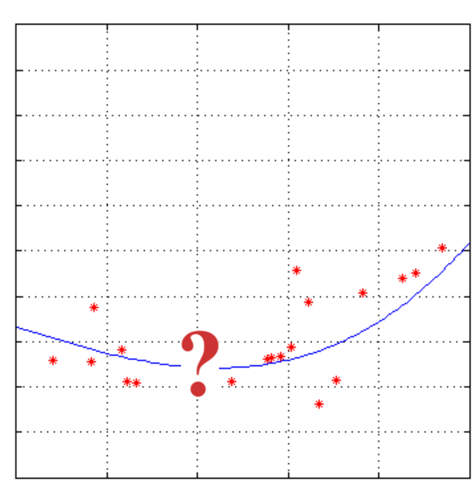


۹) یک تابع غیرخطی مناسب مانند f بر بازه‌ی $[0,1]$ و تعدادی نقطه $x_1, \cdots, x_{100}$ در این بازه انتخاب کنید. y_iها را با یک مدل رگرسیونی ساده از روی x_iها می‌سازیم: $y_i = f(x_i) + \epsilon_i$ که ϵ_iها مستقل از هم و گاوسی با میانگین صفر و واریانس σ^2 هستند.

الف) برای $D = 1,2,\cdots,10$ بهترین چندجمله‌ای $\hat{f}$ از درجه‌ی D را به داده‌ها برازش کنید.

ب) برای $D = 1,2,\cdots,10$ بهترین چندجمله‌ای مثلثاتی $\hat{f}$ از درجه‌ی D را به داده‌ها برازش کنید.

پ) با تکرار قسمت‌های الف و ب به تعداد زیاد، $\mathbb{E}[\|\hat{f} - f\|_\infty]$ را تخمین بزنید. و دو نمودار خطا بر حسب درجه، D، در هر حالت رسم کنید.



۱۰) (؟؟؟) رگرسیون چندگانه، ستیغی، لاسو، بهترین زیرمجموعه برای یک داده‌ی واقعی!