رگرسیون

۱. در مسئله رگرسیون فرض کنید

$$y = w^T \phi(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

بطوری که $y \in R,\ x \in R^d,\ \phi: R^d \to R^d,\ w \in R^{d\times 1},\ \epsilon \in R$.

نشان دهید

$$y \sim N(w^T \phi(x), \sigma^2)$$

(راهنما : درباره تابع مولد گشتاور توزیع نرمال تحقیق کنید)

۲. مسئله رگرسیون خطی ساده را در نظر بگیرید (از دیدگاه آماری).

$$y_i = w_0 + w_1 x_i + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma^2),\ \ i = 1, \dots, n$$

بنابر سوال قبل می‌توان نتیجه گرفت

$$y_i \sim N(w_0 + w_1 x_i, \sigma^2)$$

الف) در این حالت، برآورد بیشینه درست‌نمایی از پارامترهای w_0, w_1, σ^2 را بدست آورید. $(w_0^{MLE}, w_1^{MLE}, \sigma^2_{MLE})$

ب) بنابر قسمت الف، نشان دهید

$$w_0^{MLE} = \bar{y} - w_1^{MLE} . \bar{x},$$

$$w_1^{MLE} = \frac{S_{xy}}{S_x^2}$$

۳. مجموعه داده زیر را در نظر بگیرید

X	Y
5	2
0	1
2	1
1	1
2	0

بنابر برآوردگرهای سوال ۲، برآوردهای w_0, w_1 و σ^2 (واریانس خطاها) را برای این مجموعه داده بدست آورید.

۴. مسئله رگرسیون چند جمله‌ای را در نظر بگیرید بطوری که

$$y(x_n, w) = \sum_{j=0}^{M} w_j (x_n)^j \,, \quad n = 1, \dots, N \qquad (1)$$

و فرض کنید تابع خطا بصورت زیر تعریف شده است

$$E(w) = 0.5 \sum_{n=1}^{N} (y(x_n, w) - t_n)^2 \qquad (2)$$

معادله رابطه ۱ با تابع خطا رابطه ۲ را در نظر بگیرید. نشان دهید که ضرایب $w = \{w_i\}$ باید مقادیر زیر را داشته باشند تا رابطه ۲، کمینه شود.

$$\sum_{j=0}^{M} A_{ij} w_j = T_i,$$

که در آن

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j}, \qquad T_i = \sum_{n=1}^{N} (x_n)^i \, t_n$$

مقدمات

۵. مجموعه داده زیر را در نظر بگیرید

	X1	X2	X3	X4
$x_1{}^T$	2	1	1	2
$x_2{}^T$	3	0	2	4
$x_3{}^T$	4	2	3	6
$x_4{}^T$	0	1	4	8

الف) میانگین نمونه‌ها را براساس برآوردگر سازگار میانگین جامعه بدست آورید.

ب) ماتریس کواریانس نمونه را بدست آورید و هر درایه از آن را توصیف کنید.

۶. فرض کنید $X_1, \dots, X_n$ متغیرهای تصادفی مستقل و هم توزیع باشند یعنی

$$X_1, \dots, X_n \sim Ber(\theta)$$

الف) (معادل با مسئله رگرسیون لجستیک) فرض کنید نمونه‌های مشاهده شده $x_1, \dots, x_n$، به‌ترتیب از متغیرهای تصادفی بالا داریم. در این حالت، برآورد بیشینه درست‌نمایی از پارامتر θ را بدست آورید ($\hat{\theta}_{MLE}$).

ب) (معادل با مسئله رگرسیون لجستیک بیزی) در حالت الف، با استفاده از دیدگاه فراوانی، مقداری از θ را برآورد کردیم تا تابع درست‌نمایی بیشینه شود. حال فرض کنید قصد داریم از دانش قبلی برای پارامتر θ استفاده کنیم. این عمل را با قرار دادن یک توزیع بر روی θ انجام می‌دهیم. چون θ پارامتر احتمال است، بنابراین باید همیشه بین ۰ و ۱ باشد. یک توزیع مناسب برای θ، توزیع بتا است یعنی

$$\theta \sim Beta(\alpha, \beta), \qquad f(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

تابع احتمال پسین $f(\theta|x_1, \dots, x_n)$ را بدست آورید و سپس نقطه‌ای از θ را برآورد کنید تا این تابع بیشینه شود ($\hat{\theta}_{MAP}$).

$$f(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta).f(\theta)}{\int_0^1 f(x_1, \dots, x_n|\theta).f(\theta)\, d\theta}$$

۷. (امتیازی) (رگرسیون خطی بیزی) فرض کنید

$$Y|X_1, \dots, X_n, \theta \sim^{iid} N(X\theta, \sigma^2 I_n), i = 1, \dots, n$$

$$\theta \in R^{p\times 1}$$

$$Y \in R^{n\times 1}$$

ماتریس X یک ماتریس $n \times p$ است که سطر iام آن، X_i است.

الف) تابع چگالی توأم مدل بالا را بنویسید.

ب) یک توزیع پیشین مناسب ($conjugate$) برای پارامتر θ و σ^2 بصورت زیر است

$$\theta, \sigma^2 \sim N(\mu_0, \sigma^2 S_0). Inv\_Gamma(v_0, \alpha_0)$$

بطوری که $S_0 \in R^{p\times p}$ و v_0, α_0 اسکالر هستند و بر اساس اطلاعات پیشین، مقدار دقیقی را دریافت می‌کنند. (اگر هیچ اطلاعاتی موجود نباشد، از پیشین‌های فاقد اطلاعات یا non-informative استفاده می‌شود.) در این حالت توزیع پسین $p(\theta|X_1, \dots, X_n)$ را بدست آورید.