به نام خدا

پاسخ تمارین الگوریتم های یادگیری ماشین

(سری اول)

حسام مومیوند فرد

86803063

پاسخ سؤال اول

۱. در مسئله رگرسیون فرض کنید

$$y = \overbrace{w^T\phi(x)}^{x(n)} + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

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)

moment generative function of Normal Distribution is: $\exp\{t\mu + \frac{1}{2}\sigma^2 t^2\}$

$$\varepsilon \sim Normal(0, \sigma^2) \Longrightarrow M_\varepsilon(t) = \exp\{t \times 0 + \tfrac{1}{2}\sigma^2 t^2\}$$

$$\longrightarrow M_\varepsilon(t) = \exp\{\tfrac{1}{2}\sigma^2 t^2\}$$

$$M_y(t) = \mathbb{E}(\exp\{ty\}) = \mathbb{E}(\exp\{t[w^T\varphi(x) + \varepsilon]\})$$

$$= \mathbb{E}(\exp\{tw^T\varphi(x) + \varepsilon t\})$$

$$w^T\varphi(x) = \text{Constant} \Rightarrow \underbrace{\mathbb{E}(\exp\{\varepsilon t\})}_{M_\varepsilon(t)} \exp\{tw^T\varphi(x)\}$$

$$= M_\varepsilon(t) \cdot \exp\{tw^T\varphi(x)\} = \exp\{\tfrac{1}{2}\sigma^2 t^2\} \cdot \exp\{tw^T\varphi(x)\}$$

$$= \exp\{\underbrace{tw^T\varphi(x)}_{\mu'} + \tfrac{1}{2}\sigma^2 t^2\}$$: مشابه M یک توزیع نرمال

$$\rightarrow y \sim Normal(w^T\varphi(x), \sigma^2)$$

18

پاسخ سوال دوم

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

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

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

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

$$w_1^{MLE} = \frac{s_{xy}}{s_x^2}$$

الف)

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

$$\rightarrow \mathbb{P}(y_i \mid x_i, w_0, w_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - w_0 - w_1 x_i)^2}{2\sigma^2} \right\}$$

$$\Rightarrow L(w_0, w_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - w_0 - w_1 x_i)^2}{2\sigma^2} \right\}$$

$$\xrightarrow{\ln} \ln L(w_0, w_1, \sigma^2) = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \ln \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2$$

حال باید برحسب پارامترهای متفاوت w_0, w_1, σ^2 مشتق گرفته و برابر صفر قرار دهیم تا MLE به دست بیاید.

$$\frac{\partial}{\partial w_0} \ln L(w_0, w_1, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{\partial}{\partial w_0} (y_i - w_0 - w_1 x_i)^2$$

$$= -\frac{1}{\cancel{2}\sigma^2} \sum_{i=1}^{n} \cancel{2}(y_i - w_0 - w_1 x_i)\left(\overset{-1}{\cancel{\frac{\partial}{\partial w_0} -w_0}}\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial w_0} \ln L(w_0, w_1, \sigma^2) = \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i) = 0$$

$$\rightarrow \sum_{i=1}^{n} y_i = n w_0 + w_1 \sum_{i=1}^{n} x_i \quad (1)$$

$$\frac{\partial}{\partial w_1} \ln L(w_0, w_1, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{\partial}{\partial w_1} (y_i - w_0 - w_1 x_i)^2$$

$$= -\frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(y_i - w_0 - w_1 x_i)(-x_i) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i (y_i - w_0 - w_1 x_i)$$

$$\longrightarrow \sum_{i=1}^{n} x_i (y_i - w_0 - w_1 x_i) = 0 \quad ②$$

$$①, ② \Rightarrow \begin{cases} \sum_{i=1}^{n} y_i = n w_0 + w_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i (y_i - w_0 - w_1 x_i) = 0 \end{cases} \Rightarrow \begin{cases} w_0 = \bar{y} - w_1 \bar{x} \\ w_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \end{cases}$$

$$\frac{\partial}{\partial \sigma^2} \ln L(w_0, w_1, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2 = 0$$

$$\times 2\sigma^4 \Rightarrow -n\sigma^2 + \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2 = 0 \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2$$

b)

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$S_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

با جایگذاری این دو در رابطه ی w_1^{MLE} در بخش a خواهیم داشت،

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

و با جایگذاری w_1^{MLE} در w_0^{MLE} به دست آمده از بخش a خواهیم داشت

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

پاسخ سؤال سوم

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.

$$\bar{x} = \frac{5+0+2+1+2}{5} = \frac{10}{5} = 2, \qquad \bar{y} = \frac{2+1+1+1+0}{5} = \frac{5}{5} = 1$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \underbrace{\cancel{(5-2)(2-1)}}_{3} + \underbrace{\cancel{(0-2)(1-1)}}_{0} + \underbrace{\cancel{(2-2)(1-1)}}_{0} + \underbrace{\cancel{(1-2)(1-1)}}_{0} + \underbrace{\cancel{(2-2)(1-0)}}_{0} = 3$$

$$S_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \underbrace{\cancel{(5-2)^2}}_{9} + \underbrace{\cancel{(0-2)^2}}_{4} + \underbrace{\cancel{(2-2)^2}}_{0} + \underbrace{\cancel{(1-2)^2}}_{1} + \underbrace{\cancel{(2-2)^2}}_{0} = 14$$

$$\Rightarrow w_1^{MLE} = \frac{S_{xy}}{S_x^2} = \frac{3}{14}, \qquad w_0^{MLE} = \bar{y} - w_1^{MLE}\bar{x} = 1 - 2\times\frac{3}{14} = 1 - \frac{6}{14} = \frac{8}{14}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2$$

$$= \frac{1}{5}\left[\left(2 - \frac{8}{14} - \frac{3}{14}\times 5\right) + \left(1 - \frac{8}{14} - \frac{3}{14}\times 0\right) + \left(1 - \frac{8}{14} - \frac{3}{14}\times 2\right) + \left(1 - \frac{8}{14} - \frac{3}{14}\times 1\right) + \left(1 - \frac{8}{14} - \frac{3}{14}\times 2\right)\right]$$

$$\simeq 0.271$$

پاسخ سوال چهارم

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

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

هدف ما کمینه سازی خطا است پس رابطه ی ① را به ② جایگذاری کرده و بر حسب w مشتق میگیریم و برابر صفر قرار میدهیم.

$$①, ② \rightarrow E(w) = 0.5 \sum_{i=1}^{N} \left(\sum_{j=0}^{M} w_j (x_n)^j - t_n \right)^2$$

$$\frac{\partial}{\partial w_j} E(w) = \frac{1}{2} \frac{\partial}{\partial w_j} \left[\sum_{i=1}^{N} \left(\sum_{j=0}^{M} w_j (x_n)^j - t_n \right)^2 \right] = \frac{1}{2} \sum_{i=1}^{N} \frac{\partial}{\partial w_j} \Big[\underbrace{\sum_{j=0}^{M} w_j (x_n)^j - t_n}_{\alpha} \Big]^2$$

$$= \frac{1}{2} \sum_{i=1}^{N} \frac{\partial \alpha^2}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{N} 2\alpha \underbrace{\frac{\partial}{\partial w_j} \alpha}_{= (x_n)^j} = \frac{1}{2} \sum_{i=1}^{N} 2 \left(\sum_{j=0}^{M} w_j (x_n)^j - t_n \right) (x_n)^j$$

$$= \sum_{i=1}^{N} \left(\sum_{k=0}^{M} w_k (x_n)^k - t_n \right) (x_n)^j = 0$$

$$= \sum_{i=1}^{N} \left(\sum_{k=0}^{M} w_k (x_n)^{k+j} - t_n (x_n)^j \right) = 0$$

با یه مقدار جابه جا کردن طرفین به معادله ای که در صورت سوال داده شده خواهیم رسید.

پاسخ سؤال دهم

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.

a)

ماتریس داده‌ها را تشکیل می‌دهیم:

$$X = \begin{bmatrix} 2 & 1 & 1 & 2 \\ 3 & 0 & 2 & 4 \\ 4 & 2 & 3 & 6 \\ 0 & 1 & 4 & 8 \end{bmatrix}$$

$$\bar{X}_1 = \frac{2+3+4+0}{4} = \frac{9}{4}, \quad \bar{X}_2 = \frac{1+0+2+1}{4} = \frac{4}{4}, \quad \bar{X}_3 = \frac{1+2+3+4}{4} = \frac{10}{4}, \quad \bar{X}_4 = \frac{2+4+6+8}{4} = \frac{20}{4}$$

$$\Rightarrow \bar{X} = \begin{bmatrix} 2.25 \\ 1 \\ 2.5 \\ 5 \end{bmatrix}$$

b)

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})(x_i - \bar{X})^T$$

$$= \frac{1}{3} \left(\begin{bmatrix} 2-2.25 \\ 3-1 \\ 4-2.5 \\ 0-5 \end{bmatrix} \begin{bmatrix} 2-2.25 & 3-1 & 4-2.5 & 0-5 \end{bmatrix} + \begin{bmatrix} 1-2.25 \\ 0-1 \\ 2-2.5 \\ 1-5 \end{bmatrix} \begin{bmatrix} 1-2.25 & 0-1 & 2-2.5 & 1-5 \end{bmatrix} \right.$$

$$\left. + \begin{bmatrix} 1-2.25 \\ 2-1 \\ 3-2.5 \\ 4-5 \end{bmatrix} \begin{bmatrix} 1-2.25 & 2-1 & 3-2.5 & 4-5 \end{bmatrix} + \begin{bmatrix} 2-2.25 \\ 4-1 \\ 6-2.5 \\ 8-5 \end{bmatrix} \begin{bmatrix} 2-2.25 & 4-1 & 6-2.5 & 8-5 \end{bmatrix} \right)$$

$$= \frac{1}{3} \left(\begin{bmatrix} 8.75 & 1 & -2.5 & 5 \\ 1 & 2 & 1 & 2 \\ -1.5 & 1 & 5 & 10 \\ -5 & 2 & 10 & 20 \end{bmatrix} \right) = \begin{bmatrix} 2.91 & 0.33 & -0.83 & -1.66 \\ 0.33 & 0.66 & 0.33 & 0.66 \\ -0.83 & 0.33 & 1.66 & 3.33 \\ -1.66 & 0.66 & 3.33 & 6.66 \end{bmatrix}$$

در ماتریس تشکیل شده عناصر روی قطر اصلی واریانس متغیرهای مربوط به ایندکس خواهند بود و درایه‌های دیگر مربوط به کوواریانس متغیرهای
مختلف نسبت به هم هستند.

کوواریانس ←
- مثبت: به هم وابسته‌اند و در راستای هم رشد می‌کنند
- منفی: به هم وابسته‌اند اما عکس راستای هم رشد می‌کنند.
- صفر: در صورت مستقل بودن دو متغیر کوواریانس این دو صفر خواهد شد.

پاسخ سؤال ششم

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 ($\hat{\theta}_{MAP}$).

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

الف)

$$Ber(\theta)\begin{cases} 1 & \theta \\ 0 & 1-\theta \end{cases}$$

$$L(X_1=x_1, X_2=x_2, \dots, X_n=x_n|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\xrightarrow{\ln} \ln L(X_1=x_1, \dots, X_n=x_n|\theta) = \sum_{i=1}^{n} \ln\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$= \sum_{i=1}^{n} x_i \ln\theta + \sum_{i=1}^{n}(1-x_i)\ln(1-\theta) = \ln\theta\Big(\underbrace{\sum_{i=1}^{n} x_i}_{S}\Big) + \ln(1-\theta)\Big(\sum_{i=1}^{n}(1-x_i)\Big)$$

$$= S\ln\theta + (n-S)\ln(1-\theta) = \ln L(X_1=x_1, \dots, X_n=x_n|\theta)$$

$$\frac{\partial}{\partial\theta}\ln L(\dots) = 0$$

$$\longrightarrow \frac{S}{\theta} + \frac{n-S}{1-\theta} = 0 \longrightarrow \hat{\theta}_{MLE} = \frac{S}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

b) $\underset{\text{Ber}(\theta)}{L}(X_1=x_1,\dots,X_n=x_n\mid\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$

$$\sim f_{(X_1=x_1,\dots,X_n=x_n\mid\theta)} = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\sum_{i=1}^{n} x_i = S$$

$$\Rightarrow f_{(X_1=x_1,\dots,X_n=x_n\mid\theta)} = \theta^{S}(1-\theta)^{n-S}$$

قانون بیز

$$\Rightarrow f(\theta\mid X_1=x_1,\dots,X_n=x_n) \propto f_{(X_1=x_1,\dots,X_n=x_n\mid\theta)} \overbrace{f_{(\theta)}}^{\text{Beta}(\alpha,\beta)}$$

$$f_{(\theta\mid X_1=x_1,\dots,X_n=x_n)} \propto \theta^{S}(1-\theta)^{n-S}\ \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\rightarrow f_{(\theta\mid X_1=x_1,\dots,X_n=x_n)} \sim \text{Beta}(S+\alpha,\ n+\beta-S)$$