مقدمه‌ای بر یادگیری ماشین - تمرین تئوری 1

مهرانشر مظاهری - 98102346

صفحه 1

* احتمال گوسی حاشیه‌ای و شرطی

1.

(الف) • $\underline{y}$ بردار گوسی است $\Rightarrow$ هر ترکیب خطی از y_i ها گوسی خواهد بود (*)

بردار $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T$ را تعریف می کنیم؛ طبق (*)، $y_2 = e_2 \underline{y}$ هم توزیع گاوسی دارد: (1) $y_2 \sim normal$

$$E[y_2] = E[e_2 \underline{y}] \underset{\text{linearity of expectation}}{=} e_2 E[\underline{y}] = e_2 \mu = [0 \quad 1] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_2 \quad (2)$$

$$VAR[y_2] = C_{e_2 \underline{y}} = E\left[(e_2 \underline{y} - E e_2 \underline{y})(e_2 \underline{y} - E e_2 \underline{y})^T\right] \underset{\text{linearity of expectation}}{=} E\left[(e_2(\underline{y} - E\underline{y}))(e_2(\underline{y} - E\underline{y}))^T\right]$$

$$= E\left[e_2 (\underline{y} - E\underline{y})(\underline{y} - E\underline{y})^T e_2^T\right] = e_2 E\left[(\underline{y} - E\underline{y})(\underline{y} - E\underline{y})^T\right] e_2^T = e_2 C_{\underline{y}} e^T$$

$$= [0 \quad 1] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \Sigma_{22} \quad (3)$$

$$(1),(2),(3) \Rightarrow y_2 \sim normal(\mu_2, \Sigma_{22})$$

(ب)

$$y_1, y_2 \sim normal(\mu, \Sigma) \Rightarrow P(y_1 | y_2 = y_2) = \frac{P(\underline{y} = y)}{P(y_2 = y_2)} = \frac{\frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right]}{\frac{1}{\sqrt{2\pi |\Sigma_{22}|}} \exp\left[-\frac{1}{2}(y_2-\mu_2)^T \Sigma_{22}^{-1} (y_2-\mu_2)\right]}$$

الف: $y_2 \sim normal(\mu_2, \Sigma_{22})$

$P(y_2 = y_2)$ به ازای هر مقدار مشخص y_2 مقدار ثابت است (مخرج) و $P(\underline{y} = y)$ نیز توزیع نرمال دارد، پس

$P(y_1 | y_2) = \frac{\text{توزیع نرمال وابسته به } y_2}{\text{عدد ثابت وابسته به } y_2}$ و $y_1 | y_2 = y_2$ توزیع گاوسی دارد: (1) $y_1 | y_2 \sim normal$

متغیر تصادفی $z = x_1 - \Sigma_{12} \Sigma_{22}^{-1} x_2$ را تعریف می کنیم:

$$cov(z, y_2) = cov(y_1, y_2) - \Sigma_{12} \Sigma_{22}^{-1} cov(y_2, y_2) = \Sigma_{12} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22} = 0 \Rightarrow$$

$$z \,\&\, y_2 \text{ are uncorrelated} \xrightarrow{\underline{y} \sim N} z \,\&\, y_2 \text{ are independent } **$$

$$E(y_1 | y_2) = E(z + \Sigma_{12} \Sigma_{22} y_2 | y_2) = E(z | y_2) + \Sigma_{12} \Sigma_{22}^{-1} E(y_2 | y_2) \overset{**}{=} E(z) + \Sigma_{12} \Sigma_{22}^{-1} y_2$$

$$= \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} (\mu_2 - y_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (y_2 - \mu_2) \quad (2)$$

$$VAR(y_1 | y_2) = VAR(z + \Sigma_{12} \Sigma_{22} y_2 | y_2) = VAR(z | y_2) + \cancel{VAR(\Sigma_{12} \Sigma_{22} y_2 | y_2)}^{0} - \cancel{2cov(z, \Sigma_{12} \Sigma_{22}^{-1} y_2 | y_2)}^{0}$$

$$= VAR(z | y_2) \overset{**}{=} VAR(z) = VAR(x_1 - \Sigma_{12} \Sigma_{22}^{-1} x_2) = VAR(x_1) + (\Sigma_{12} \Sigma_{22}^{-1})^2 VAR(x_2) + 2cov(x_1, -\Sigma_{12} \Sigma_{22}^{-1} x_2)$$

صفحه 2

این تعریف متغیر z از یکی از فوروم‌های اینترنتی گرفته شده

$$= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (3)$$

$$(1)\&(2)\&(3): \quad y_1 | y_2 \sim normal\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

۲. (الف) بردار تصادفی y را به شکل روبرو می‌نویسیم : $\underline{y} = W\underline{z} + \underline{U}, \quad \underline{U} \sim normal(\underline{b}, \Sigma_y)$

$$E\,\underline{y} = E[E[\underline{y}|\underline{z}]] = E[W\underline{z} + \underline{b}] = W E[\underline{z}] + \underline{b} = W\mu_z + \underline{b} \quad (1)$$

در تشکیل فرم $\underline{y}$، بردار تصادفی $\underline{U}$ عمود بر $\underline{z}$ تعریف می‌شود ($cov(\underline{U},\underline{z}) = 0$ (*)) چرا که طبق فرض سوال $\Sigma_{\underline{y}|\underline{z}} = \Sigma_{\underline{y}}$ است.

$$Cov(\underline{y}, \underline{z}) = cov(W\underline{z} + \underline{U}, \underline{z}) \rightarrow = W\,cov(\underline{z},\underline{z}) + cov(\underline{U},\underline{z})$$

$$\overset{(*)}{=} W\,\Sigma_z + 0 \quad (2), \quad Similarly: cov(\underline{z},\underline{y}) = \Sigma_z W^T \quad (3)$$

$$cov(\underline{y},\underline{y}) = cov(W\underline{z} + \underline{U}, W\underline{z} + \underline{U}) = W\,\overset{\Sigma_z}{\cancel{cov(\underline{z},\underline{z})}}\,W^T + W\,cov(\underline{z},\underline{U}) + cov(\underline{z},\underline{U})W^T$$

$$+ cov(\underline{U},\underline{U}) \overset{(*)}{=} W\Sigma_z W^T + \Sigma_{\underline{y}} \quad (4)$$

$$(1)\&(2)\&(3)\&(4): \quad P(z,y) = normal\left(\begin{bmatrix}\mu_z \\ W\mu_z + b\end{bmatrix}, \begin{bmatrix}\Sigma_z & \Sigma_z W^T \\ W\Sigma_z & W\Sigma_z W^T + \Sigma_y\end{bmatrix}\right)$$

ب) $P(y,z)$ را در بخش الف بدست آوردیم. محاسبه $P(x_1|x_2)$ از روی $P(x_1, x_2)$ را برای بردار تصادفی نرمال بدست آوردیم (در بخش ب سوال 1)

$$P(y_1|y_2) = N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

$$P(z|y) = N(\mu_{z|y}, \Sigma_{z|y}), \quad \mu_{z|y} = \mu_z + \Sigma_z W^T (W\Sigma_z W^T + \Sigma_y)^{-1}(y - \overset{b}{\cancel{\mu_y}}) \quad (1)$$

$$\Sigma_{z|y} = \Sigma_z - \Sigma_z W^T (W\Sigma_z W^T + \Sigma_y)^{-1} W\Sigma_z = \Sigma_z(I - W^T(W\Sigma_z W^T + \Sigma_y)^{-1}W\Sigma_z) \rightarrow$$

$$\Sigma_{z|y}^{-1} = \Sigma_z^{-1} + W^T\Sigma_y^{-1}W \qquad (1): \mu_{z|y} = \Sigma_{z|y}\left(\Sigma_z^{-1}\mu_z + W^T\Sigma_y^{-1}(y-b) + W^T\Sigma_y^{-1}W\mu_z\right)$$

$$\rightarrow \Sigma_{z|y}\left(\Sigma_z^{-1}\mu_z + W^T\Sigma_y^{-1}(y-b)\right)$$

۲ محاسبه احتمال

۱.

X, Y are jointly normal $\Rightarrow X+Y \sim normal \rightarrow X+Y \sim N(-1,3)$

$E[X+Y] = \mu_X + \mu_Y = 0 + (-1) = -1$

$VAR[X+Y] = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 1 + 4 + (-2) = 3$

$P(X+Y>0) = 0.281 = \Phi(\frac{1}{\sqrt{3}}) + 1$

۲.

هر دو متغیر مشترکا نرمال هستند $\Leftarrow$ [استقلال $\Leftrightarrow$ غیر همبستگی $[cov(,)=0]$]

$0 = cov(X+2Y, aX+Y) = a\, \overset{1}{cov(X,X)} + (2a+1)\, \overset{-1}{cov(X,Y)} + 2\, \overset{4}{cov(Y,Y)}$

$\Rightarrow -a - 1 + 8 = 0 \Rightarrow \boxed{a = 7}$

۳.

$X+Y \sim normal(-1,3)$

$2X - Y \sim normal(1, \overset{12}{4\times 1 + 4 + 4})$ ، $2\mu_X - \mu_Y$

$cov(X+Y, 2X-Y) = 2\sigma_X^2 - \sigma_Y^2 + (-1) = -3$

$\rho = \frac{-3}{12} = -\frac{1}{4}$

$X+Y \mid 2X-Y=0 \sim N(-1 + \frac{1}{4} \times \frac{-\sqrt{3}}{\sqrt{12}}, \frac{15}{16}\times 3)$

$E[Y \mid X=\alpha] = \mu_Y + \rho\sigma_Y \frac{\alpha - \mu_X}{\sigma_X}$

$VAR[Y \mid X=\alpha] = (1-\rho^2)\sigma_Y^2$

$N(-0.875, \frac{45}{16})$

$P(X+Y>0 \mid 2X-Y=0) = 0.3 = \Phi(\frac{0.875}{\sqrt{\frac{45}{16}}})$

۳ مقادیر ویژه

ماتریس A را با مقادیر ویژه اش تجزیه می کنیم $(A = P\Lambda P^{-1}, \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix})$:

$$e^A = e^{P\Lambda P^{-1}} = P e^{\Lambda} P^{-1} \overset{(*)}{=} P \times \begin{bmatrix} \frac{\lambda_1 e^{\lambda_1} + \lambda_2 e^{\lambda_2} + (-\lambda_2 e^{\lambda_2}) - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} & 0 \\ 0 & \frac{\lambda_1 e^{\lambda_2} + \lambda_2 e^{\lambda_1} - \lambda_1 e^{\lambda_1} - \lambda_2 e^{\lambda_2}}{\lambda_1 - \lambda_2} \end{bmatrix} \times P^{-1}$$

$$= P \times \left(\begin{bmatrix} \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} & 0 \\ 0 & \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} \end{bmatrix} + \begin{bmatrix} \frac{\lambda_1 e^{\lambda_1} - \lambda_1 e^{\lambda_2}}{\lambda_1 - \lambda_2} & 0 \\ 0 & \frac{\lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_2}}{\lambda_1 - \lambda_2} \end{bmatrix} \right) P^{-1}$$

$$= \underbrace{P I P^{-1}}_{I} \left(\frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} \right) + \frac{e^{\lambda_1} - e^{\lambda_2}}{\lambda_1 - \lambda_2} \times \underbrace{P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}}_{A} = \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{e^{\lambda_1} - e^{\lambda_2}}{\lambda_1 - \lambda_2} A$$

$(*)$: در این مرحله $\begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}$ ، با اضافه و کم کردن جملات و هم ضرب و تقسیم کردن $(\lambda_1 - \lambda_2)$ ، ماتریس معادل e^{Λ} را در ادامه قرار داده ایم تا با فاکتورگیری و تفکیک به عبارت مدنظر برسیم

MAP ۴

۱.

X, Y are independent gaussian RVs. : $f_{X,Y}(x,y) \overset{\hat{\mu}=\mu \text{ با}}{=} f_X(x|\mu=\hat{\mu})\, f_Y(y|\mu=\hat{\mu}) = \emptyset$ فرض سوال

$$f_{X,Y,\mu}(x,y,\mu) = f_{X,Y}(x,y|\mu)\, f_M(\mu) = f_X(x|\mu)\, f_Y(y|\mu)\, f_\mu(t)$$

$$f_{X,Y,\mu}(x,y,t) = \begin{cases} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x-t)^2 - \frac{1}{2}(y-t)^2\right) & 0<t<1 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\mu}_{MAP} = \underset{\mu}{\operatorname{argmax}} \begin{cases} \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left[(x-\mu)^2+(y-\mu)^2\right]\right) & 0<\mu<1 \\ 0 & OW \end{cases} \xrightarrow[\times(-2)]{\log} \underset{min}{\operatorname{arg}\cancel{max}} \left((x-\mu)^2+(y-\mu)^2\right)$$ ۲

$$\hat{\mu}_{MAP} = \underset{\mu}{\operatorname{argmin}} \left((x-\mu)^2+(y-\mu)^2\right) \longrightarrow \frac{\partial}{\partial \hat{\mu}}\left((x-\hat{\mu})^2+(y-\hat{\mu})^2\right)\Big|_{\mu=\hat{\mu}_{MAP}} = 0 \Rightarrow \boxed{\hat{\mu}_{MAP} = \frac{x+y}{2}}$$

اگر $\frac{x+y}{2}$ در تخمین بالا خارج از $[0,1]$ قرار گرفت به این معنی است نقطه بهینه برای تخمین $\hat{\mu}_{MAP}$ دیگر روی نقطه بحرانی ($\frac{\partial}{\partial \mu} = 0$) نخواهد بود بلکه برابر با 1 یا 0 (هر کدام به $\frac{x+y}{2}$ نزدیکتر بود) خواهد بود (این اثر prior بر روی جواب است)

$$\Rightarrow \hat{\mu}_{MAP} = \begin{cases} \frac{x+y}{2} & 0 \leqslant \frac{x+y}{2} \leqslant 1 \\ 1 & 1 < \frac{x+y}{2} \\ 0 & \frac{x+y}{2} < 0 \end{cases}$$

RHS: $E(VAR(X|Y)) + VAR(E(X|Y))$

۵. متغیر تصادفی کمکی

در بسط رابطه بالا از $(*) VAR(X) = EX^2 - E^2X$, $(**) E[E[X|Y]] = 0$ استفاده می کنیم

$$\overset{(*)}{=} E\left(E[X^2|Y] - \cancel{E^2[X|Y]}\right) + \cancel{E(E^2[X|Y])} - E^2(E[X|Y]) = E[E[X^2|Y]] - E^2[E[X|Y]] \overset{(**)}{=}$$

$$EX^2 - E^2X \overset{(*)}{=} VAR[X]$$