HW01_machine learning

Saeed [illegible]

Student ID: 98106542

---

① conditional Bivariate normall distribution

suppose $\mathbf{y} \sim \mathcal{N}(\mu, \Sigma)$, $\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, Prove that:

a) $p(\mathbf{y_2}) = \mathcal{N}(\mu_2, \mathbf{\Sigma_{22}})$

b) $p(\mathbf{y_1}|\mathbf{y_2}) = \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y_2} - \mu_2), \mathbf{\Sigma_{11}\Sigma_{12}\Sigma_{22}}^{-1}\mathbf{\Sigma_{21}})$

a) If we define $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we know that $y_2 = [0, 1]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = [0, 1]Y$

$$E(y_2) = E(e_2^T Y) = e_2^T E(Y) = [0,1]\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_2 \quad (1)$$

$$var(y_2) = E\left([y_2 - Ey_2][y_2 - Ey_2]\right) = E\left([e_2^T Y - e_2^T E(Y)][e_2^T Y - e_2^T E(Y)]^T\right)$$

$$E\left[e_2^T (Y - E(Y))(Y - E(Y))^T e_2\right] = e_2^T E\left([Y - E(Y)][Y - E(Y)]^T\right) e_2$$

$$\Rightarrow [0 \ 1]\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \Sigma_{22} \quad (2)$$

(3) we know that each element of Y is a gaussian random variable so $e_2^T Y = y_2$ has a gaussian distribution

$$(1)(2)(3) \longrightarrow \boxed{y_2 \sim N(\mu_2, \Sigma_{22})}$$

b) $p(\mathbf{y_1}|\mathbf{y_2}) = \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y_2} - \mu_2), \boldsymbol{\Sigma_{11}\Sigma_{12}\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}})$ we define two new variable:

$Z = y_1 + Ay_2$, where $A = -\Sigma_{12}\Sigma_{22}^{-1}$

first we show that Z and y_2 are independent. to do this we compute $\operatorname{cov}(Z, y_2)$

$$\rightarrow \operatorname{cov}(Z, y_2) = \operatorname{cov}(y_1 + Ay_2, y_2) = \underbrace{\operatorname{cov}(y_1, y_2)}_{\Sigma_{12}} + \underbrace{A}_{-\Sigma_{12}\cancel{\Sigma_{22}^{-1}}\cancel{\Sigma_{22}} \; I}\operatorname{cov}(y_2, y_2) = 0 \quad (1)$$

$\xrightarrow{(1)}$ (2) Because those variable are jointly normal

$$\boxed{Z, y_2 \text{ are independent}} \quad *$$

first, we compute $E(Z)$

$$\boxed{\rightarrow E(Z) = E(y_1 + Ay_2) \xrightarrow{\text{Linearity}} \mu_1 + A\mu_2} \quad **$$

now, we compute $E(y_1|y_2) \xrightarrow{y_1 = Z - Ay_2} E(y_1|y_2) = E(Z - Ay_2|y_2)$

$$\xrightarrow{\text{linearity}} \underbrace{E(Z|y_2)}_{*\ \text{independent}} - E(Ay_2|y_2) = E(Z) - A\underbrace{E(y_2|y_2)}_{y_2}$$

$$\overset{**}{\Longrightarrow} \mu_1 + A\mu_2 - Ay_2 = \mu_1 + A(\mu_2 - y_2) \Rightarrow \boxed{E(y_1|y_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)} \quad (I)$$

now, we compute $\operatorname{var}(y_1|y_2) = \operatorname{var}(Z - Ay_2|y_2)$

$$\underbrace{\operatorname{var}(Z)}_{} - \underbrace{\cancel{\operatorname{var}(Ay_2|y_2)}}_{0} \Leftarrow$$ دو متغیر مستقل هستند $\underbrace{Ay_2|y_2}_{0}$ ، $\underbrace{Z|y_2}_{\sim Z}$

$\rightarrow$ we have to compute $\operatorname{var}(Z)$

$$var(Z) = Var(y_1 + Ay_2) = var(y_1) + \underbrace{A^2 Var(y_2)}_{A^2 \Sigma_{22}} + \underbrace{2cov(y_1, Ay_2)}_{2A\Sigma_{12}}$$

$var(y_1) \to \Sigma_{11}$

$2A\Sigma_{12} \to -2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}$

$$A^2 = AA^T = (-\Sigma_{12}\Sigma_{22}^{-1})(-\Sigma_{12}\Sigma_{22}^{-1})\underbrace{\Sigma_{22}}_{I}$$

$$\Rightarrow var(Z) = \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}$$

$$\Rightarrow \boxed{var(y_1|y_2) = var(Z) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}} \quad \text{II}$$

$$\text{I} \ \text{II} \longrightarrow \boxed{y_1|y_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12})}$$

1.2

suppose

$$p(z) = \mathcal{N}(\mu_z, \Sigma_z)$$
$$p(y|z) = \mathcal{N}(Wz + b, \Sigma_y)$$

. prove that:

$$p(z, y) = \mathcal{N}(\mu, \mathbf{\Sigma})$$
$$\mu = \begin{bmatrix} \mu_z \\ W\mu_y + b \end{bmatrix}$$
$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_z & \Sigma_z W^T \\ W\Sigma_z & \Sigma_y + W\Sigma_z W^T \end{bmatrix}$$

we use properties of previous section

we know that $y_1|y_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} \ldots)$

مقایسه

$$\Longrightarrow Wz + b \triangleq \Sigma_{yz}\Sigma_z^{-1}(z - \mu_z) + \mu_y,$$

$$\Rightarrow \begin{cases} (1)\ W = \Sigma_{yz}\Sigma_z^{-1} \\ (2)\ b = \mu_y - \Sigma_{yz}\Sigma_z^{-1}\mu_z \xrightarrow{(1)} b = \mu_y - W\mu_z \end{cases}$$

$$\Rightarrow \boxed{\mu_y = W\mu_z + b} \ *$$

$$\xrightarrow{(1)} W = \Sigma_{y_2}\Sigma_{22}^{-1} \Rightarrow \boxed{\Sigma_{y2} = W\Sigma_{22} = \Sigma_z W^T} \ **$$

داریم

$$\Rightarrow \Sigma_y = \Sigma_y - \cancel{\Sigma_{12}\Sigma_2^{-1}\Sigma_{21}} \Rightarrow$$

$$\Sigma_{y2}\Sigma_2^{-1}\Sigma_{2y} = W\Sigma_2 W^T$$

$$\Sigma_{y2}\Sigma_2^{-1}\underbrace{\cancel{\Sigma_2\Sigma_2^{-1}}}_{I}\Sigma_{y2}^T$$

$$\Rightarrow \boxed{\Sigma_y = \Sigma_{y2}\Sigma_2^{-1}\Sigma_{y2}} \quad ***$$

$$\xrightarrow{*} \quad \xrightarrow{**} \quad \xrightarrow{***}$$

حکم اثبات شد:

2.2

prove that

$$p(z|y) = \mathcal{N}(\mu_{z|y}, \Sigma_{z|y})$$
$$\mu_{z|y} = \Sigma_{z|y}[W^T\Sigma_y^{-1}(y-b) + \Sigma^{-1}\mu_z]$$
$$\Sigma_{z|y}^{-1} = \Sigma_z^{-1} + W^T\Sigma_y^{-1}W$$

از بخش 1 می‌دانیم که:

$$p(z|y) = \mathcal{N}\left(\mu_z + \Sigma_{zy}\Sigma_y^{-1}(y-\mu_y),\ \Sigma_z - \Sigma_{zy}\Sigma_y^{-1}\Sigma_{zy}\right) \quad (1)$$

از بخش 2 می‌دانیم:

$$\begin{cases} \Sigma_{yz} = W\Sigma_z \\ \mu_y = W\mu_z + b \\ \Sigma_{y|z} = \Sigma_y \end{cases} \quad (2)$$

$$\xrightarrow{(1)(2)} \mu_{z|y} = \mu_z + \Sigma_z W^T\Sigma_y^{-1}[\underbrace{y - W\mu_z - b}_{(y-b)-W\mu_z}]$$

$$\Rightarrow \mu_{z|y} = \mu_z + \Sigma_z W^T\Sigma_y^{-1}[y-b] - \Sigma_z W^T\Sigma_y^{-1}W\mu_z$$

$$\Rightarrow \boxed{\mu_{z|y} = \mu_z\left[I - \Sigma_z W^T\Sigma_y^{-1}W\right] + \Sigma_z W^T\Sigma_y^{-1}[y-b]}$$

#

①② $\longrightarrow \Sigma_{z|y} = \Sigma_z - \Sigma_z w^T \Sigma_y^{-1} w \Sigma_z = \Sigma_z \left(I - w^T \Sigma_y^{-1} w \Sigma_z \right)$

$$\Rightarrow \boxed{\Sigma_z^{-1} \Sigma_{z|y} = I - w^T \Sigma_y^{-1} w \Sigma_z} \quad **$$

← حال که رابطه‌ای برای $I - w^T \Sigma_y^{-1} w \Sigma_z$ بدست آوردیم آن را در رابطه‌ی * جایگذاری می‌کنیم:

$$\Rightarrow \boxed{\mu_{z|y} = \mu_z \left[\underbrace{I - \Sigma_z w^T \Sigma_y^{-1} w}_{\Sigma_z^{-1}\Sigma_{z|y}}\right] + \Sigma_z w^T \Sigma_y^{-1} [y - b]}$$

$$\Rightarrow \mu_{z|y} = \mu_z \Sigma_z^{-1} \Sigma_{z|y} + \Sigma_z w^T \Sigma_y^{-1} (y - b) \xrightarrow[\Sigma_{z|y} \,\cdot\, \Sigma_z]{\text{طبق مفروضات مسئله}}$$

$$\Rightarrow \mu_{z|y} = \mu_z \Sigma_z^{-1} \Sigma_{z|y} + \Sigma_{z|y} w^T \Sigma_y^{-1} (y - b)$$

$$\xrightarrow{\text{فاکتورگیری از اکطر}} \boxed{\Sigma_{z|y} \left[\Sigma_z^{-1} \mu_z + w^T \Sigma_y^{-1} (y - b) \right] = \mu_{z|y}} \rightarrow \text{میانگین}$$

$$\xrightarrow{**} I - w^T \Sigma_y^{-1} w \Sigma_z = \Sigma_z^{-1} \Sigma_{z|y} \xrightarrow[\Sigma_{z|y} \,\cdot\, \Sigma_z]{\text{طبق مفروضات مسئله}}$$

$$I - w^T \Sigma_y^{-1} w \Sigma_z = I$$

$$\xrightarrow{\times \Sigma_{z|y}^{-1}} \Sigma_{z|y}^{-1} - w^T \Sigma_y^{-1} w \overbrace{\cancel{\Sigma_z \Sigma_z^{-1}}}^{I} = \Sigma_z^{-1}$$

$$\Rightarrow \boxed{\Sigma_{z|y}^{-1} = \Sigma_z^{-1} + w^T \Sigma_y^{-1} w} \rightarrow \checkmark \quad \text{Yes!}$$

(2) Bivarite normal distribution

$$\rho = -\frac{1}{2}, \begin{cases} \mu_x = 0 \\ \sigma_x^2 = 1 \end{cases}, \begin{cases} \mu_y = -1 \\ \sigma_y^2 = 4 \end{cases}$$

a) $P(X+Y>0)$

we know that x and y are Jointly normall. so each linear combination of these two random variable is normall too.

$$E(X+Y) \xrightarrow{\text{linearity}} E(X)+E(Y) = -1$$

$$Var(X+Y) = \underbrace{Var(X)}_{1} + \underbrace{Var(Y)}_{4} + \underbrace{2cov(X,Y)}_{} = +3$$

$$\rho = \frac{cov(X,Y)}{\sigma_x \sigma_y} \xrightarrow[\substack{\sigma_x = 0 \\ \sigma_y = 2}]{\rho = -\frac{1}{2}} cov(X,Y) = -1$$

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

$$\xrightarrow[F_X(x) = P(X<x) = \Phi\left(\frac{X-\mu}{\sigma}\right)]{\text{we know that for } X \sim N(\mu, \sigma^2)} P(X+Y>0) = 1 - \Phi\left(\frac{X+1}{\sqrt{3}}\right) \simeq \boxed{0.2819}$$

b) (1) we know that $aX+Y$ and $X+2Y$ are independent

(2) we also know that $aX+Y$ and $X+2Y$ are normall.

$$\xrightarrow{(1)} cov(aX+Y, X+2Y) = 0$$

$$\Rightarrow \underbrace{a\,cov(X,X)}_{a} + \underbrace{2a\,cov(X,Y)}_{(-2a)} + \underbrace{cov(Y,X)}_{(-1)} + \underbrace{2cov(Y,Y)}_{8} = 0$$

$$a + (-2a) + (-1) + 8 = 0 \Rightarrow \boxed{a = 7}$$

c) $P(X+Y \mid 2X-Y=0)$

before solving this problem, I use the theorem that I explain in the following:

theorem:

$\boxed{\begin{array}{l}\text{suppose } X \text{ and } Y \text{ are jointly normal with parameters } \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho \\ \text{then given } X=x,\ Y \text{ is normally distributed with} \\ E(Y \mid X=x) = \mu_y + \rho\sigma_y \cdot \left(\frac{x-\mu_x}{\sigma_x}\right) \\ var(Y \mid X=x) = (1-\rho^2)\sigma_Y^2\end{array}}$

$\rightarrow E(2X-Y) = 2E(X) - E(Y) = \underline{1}$

$var(2X-Y) = \underbrace{var(2X)}_{4\,var(X)} + var(-Y) + 2cov(2X,-Y)$

$4\,var(X) + var(Y) - 4cov(X,Y)$

$\Rightarrow 4+4-4(-1) = 12 \Rightarrow \boxed{\begin{cases}\mu_{2X-Y} = \underline{1} \\ var(2X-Y) = \underline{\underline{12}}\end{cases}}$

$cov(X+Y, 2X-Y) = \underbrace{2cov(X,X)}_{2} + \underbrace{cov(X,Y)}_{-1} - \underbrace{cov(Y,Y)}_{-4} = -3$

$\boxed{\rho = \frac{-3}{\sqrt{12}\times\sqrt{3}} = -\frac{1}{2}}$

$\Rightarrow E(X+Y \mid 2X+Y=0) \xRightarrow{\text{theorem}} -1 + \left(-\frac{1}{2}\right)(\sqrt{3})\left(\frac{-1}{\sqrt{12}}\right) = -3/4$

$var(X+Y \mid 2X+Y=0) \xRightarrow{\text{theorem}} \frac{3}{4} \times 3 = \frac{9}{4}$

$\Rightarrow \boxed{P(X+Y \mid 2X+Y=0) \sim N\left(-\frac{3}{4}, \frac{9}{4}\right)}$

③ eigenvalues

$$e^A = \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{e^{\lambda_1} - e^{\lambda_2}}{\lambda_1 - \lambda_2} A$$

from eigenvalue decomposition we know that for an arbitrary matrix A we have $A = P \Lambda P^{-1}$, where Λ is diagonal matrix with eigenvalues of A on it's diagonals!

also we know that $e^{P\Lambda P^{-1}} = P e^{\Lambda} P^{-1} \Rightarrow e^A = P \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix} P^{-1}$

good to mention that:

$$e^{\lambda_1} = \frac{\lambda_1 e^{\lambda_1} + \lambda_1 e^{\lambda_2} - \lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2}$$

$$\Rightarrow P \begin{bmatrix} \frac{\lambda_1 e^{\lambda_1} + \lambda_1 e^{\lambda_2} - \lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} & 0 \\ 0 & \frac{\lambda_1 e^{\lambda_2} + \lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_2}}{\lambda_1 - \lambda_2} \end{bmatrix} P^{-1}$$

$$\rightarrow P \Bigg(\underbrace{\left[I \times \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} \right]}_{*1} + \underbrace{\begin{bmatrix} \lambda_1 e^{\lambda_1} - \lambda_1 e^{\lambda_2} & 0 \\ 0 & \lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_2} \end{bmatrix} \frac{1}{\lambda_1 - \lambda_2}}_{*2} \Bigg) P^{-1}$$

$$\frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} [P I P^{-1}] + \frac{P}{\lambda_1 - \lambda_2} \begin{bmatrix} \lambda_1 e^{\lambda_1} - \lambda_1 e^{\lambda_2} & 0 \\ 0 & \lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_2} \end{bmatrix} P^{-1}$$

$$\Rightarrow \underbrace{\frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2}\left[\overset{I}{\cancel{P I P^{-1}}}\right]}_{*1} + \underbrace{\frac{P}{\lambda_1 - \lambda_2}\begin{bmatrix} \lambda_1 e^{\lambda_1} - \lambda_1 e^{\lambda_2} & 0 \\ 0 & \lambda_2 e^{\lambda_1} - \lambda_2 e^{\lambda_2} \end{bmatrix} P^{-1}}_{*2}$$

$$*1 \rightarrow \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} e^{\lambda_1} - e^{\lambda_2} & 0 \\ 0 & e^{\lambda_1} - e^{\lambda_2} \end{bmatrix} P^{-1}}{\lambda_1 - \lambda_2}$$

$$\rightarrow \frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{\overset{A}{\cancel{P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}}}\left[e^{\lambda_1} - e^{\lambda_2}\right]}{\lambda_1 - \lambda_2}$$

$$\Rightarrow \boxed{\frac{\lambda_1 e^{\lambda_2} - \lambda_2 e^{\lambda_1}}{\lambda_1 - \lambda_2} I + \frac{e^{\lambda_1} - e^{\lambda_2}}{\lambda_1 - \lambda_2} A}$$

4 MAP $\quad \mu \sim uni(0,1)$ $\quad$ x, y independent variable

Variance $\triangleq 1$ $\quad$ guassian

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-[x-\mu_x]^2}{2}\right)}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-(y-\mu_y)^2}{2}\right)}$$

$$\Rightarrow f_{X,Y}(x,y) \xrightarrow{\text{indepednt}} f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

$$\Rightarrow f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\left[\frac{(x-\mu_x)^2}{2} + \frac{(y-\mu_y)^2}{2}\right]}$$

$$\text{MAP: minimize } f_{y|x} f_x \xrightarrow[f_{y|x} \triangleq f(y)]{} \text{minimize } f_Y f_X \triangleq \underset{\substack{X=x \\ Y=y}}{\text{minimize}} \ \frac{1}{2\pi} e^{-\left[\frac{(x-\mu_x)^2}{2} + \frac{(y-\mu_y)^2}{2}\right]}$$

$$\frac{\partial}{\partial \mu_x}(g(x)) = 0 \Rightarrow \left[+2\left(-\frac{1}{2}\right)(x-\mu_x)\right] = 0 \Rightarrow \boxed{\mu_x = x}$$

$$\frac{\partial}{\partial \mu_y}(g(x)) = 0 = [\ \text{"} \quad \text{"} \quad \text{"}\ ] = 0 \Rightarrow \boxed{\mu_y = y}$$

⑤ prove following equation :

$$Var(X) = E(Var(X|y)) + Var(E(X|y))$$

We assume $V = Var(X|y)$ and $Z = E(X|y) = \mu_{X|y}$

$$\Rightarrow Var(X) = \underbrace{E(V)}_{*} + \underbrace{Var(Z)}_{**}$$

ابتدا رابطه‌ای برای $Var(X|y)$ پیدا کنم:

$$\boxed{V = Var(X|y)} = E\left([X - \mu_{X|y}(y)]^2 \mid Y = y\right)$$

$$\Rightarrow \sum_{x_i \in R_X} (x - \mu_{X|y}(y))^2 P_{X|y}(x_i) = \boxed{E[X^2 \mid Y = y] - \mu^2_{X|y}(y)}$$

$$\xrightarrow{*} E(V) = E\left(E(X^2 \mid Y = y) - \mu^2_{X|y}(y)\right) \xrightarrow{\text{linearity}}$$

$$\boxed{E(V) = E(X^2) - E(Z^2)} \quad \text{equation } 1$$

$$\xrightarrow{**} Var(Z) = E(Z^2) - E^2(Z)$$

$$Z = E(X|y) \longrightarrow \boxed{Var(Z) = E(Z^2) - E^2(X)}$$

equation 2

$$\xrightarrow{eq\,1 + eq\,2} Var(X) = E(X^2) - \cancel{E(Z^2)} + \cancel{E(Z^2)} - E^2(X)$$

$$\Rightarrow \boxed{Var(X) = E(X^2) - E^2(X)}$$

it's definition of variance

اثبات شد :)