1.1) Q1.(a)

PIYUSHI MANUPRIYA
CS5590 '22 TA member

Case: X & Y both continuous

Starting with the conditional CDF:

$$F_{X|Y}(x|y) = \lim_{\Delta y \downarrow 0} P[X \le x \mid y \le Y \le y+\Delta y]$$

$$= \lim_{\Delta y \downarrow 0} \frac{P[X \le x \mid y \le Y \le y+\Delta y]}{P[y \le Y \le y+\Delta y]}$$

$$= \lim_{\Delta y \downarrow 0} \frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta y} f_{XY}(x', y')\, dy'\, dx'}{\int_{y}^{y+\Delta y} f_Y(y')\, dy'}$$

$$= \lim_{\Delta y \downarrow 0} \frac{\int_{-\infty}^{x} f_{XY}(x', y)\, \Delta y \, dx'}{f_Y(y)\, \Delta y}$$

where we have used that on a small interval, the density function is almost constant.

$$\Rightarrow F_{X|Y}(x|y) = \int_{-\infty}^{x} \frac{f_{XY}(x', y)}{f_Y(y)}\, dx'$$

$\therefore$ $\dfrac{f_{XY}(x, y)}{f_Y(y)}$ must be the conditional P.D.F. $f_{X|Y}(x|y)$ $\forall\, y$ s.t. $f_Y(y) \neq 0$

Hence, proved that $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$

$\forall\, y$ s.t. $f_Y(y) \neq 0$

Bayes theorem.

Case: X continuous Y discrete

$$f_{Y|X}(y|x) = \lim_{\Delta x \downarrow 0} P[Y=y \mid X \in [x, x+\Delta x]]$$

$$= \lim_{\Delta x \downarrow 0} \frac{P[Y=y,\, x \leq X \leq x+\Delta x]}{P[x \leq X \leq x+\Delta x]}$$

$$= \lim_{\Delta x \downarrow 0} \frac{P[x \leq X \leq x+\Delta x \mid Y=y]\, P[Y=y]}{P[x \leq X \leq x+\Delta x]}$$

$$= \lim_{\Delta x \downarrow 0} \frac{\int_x^{x+\Delta x} f_{X|Y}(x'|y)\, dx'\, f_Y(y)}{\int_x^{x+\Delta x} f_X(x')\, dx'}$$

$$= \lim_{\Delta x \downarrow 0} \frac{f_{X|Y}(x|y)\, \Delta x\, f_Y(y)}{f_X(x)\, \Delta x}$$

where we have used that on a small interval, the density function is almost constant.

$$\Rightarrow f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)}$$

Bayes theorem proved.

Note that $f_{Y|X}(\cdot)$ & $f_Y(\cdot)$ are PMFs whereas $f_{X|Y}(\cdot)$ is a PDF

Case: X & Y are discrete with joint PMF, f_{XY}

By definition of conditional probability over events,

$$P[X=x \mid Y=y] = \frac{P[X=x, Y=y]}{P[Y=y]}$$

$$= \frac{f_{XY}(x, y)}{f_Y(y)}$$

we denote conditional PMF of X given $Y=y$ by $f_{X|Y}(x|y)$

Hence, $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$

Q1 (b) Case : X & Y are continuous
(other cases can be done similarly)

<u>Notations</u> $g(x,y)$: joint PDF of X & Y

Marginal $g(x)$: $\int g(x,y)\,dy$

Conditional $g(y|x)$: $\dfrac{g(x,y)}{g(x)}$

$$LHS = E[f(X,Y)] = \iint f(x,y)\, g(x,y)\, dx\, dy$$

$$E_X[E_{Y|X}[f(X,Y)|X]]$$

$$= \int E_{Y|x}[f(x,Y)|x]\, g(x)\, dx$$

$$= \iint f(x,y)\, g(y|x)\, g(x)\, dy\, dx$$

$$= \iint f(x,y)\, g(x,y)\, dx\, dy$$ using Bayes Theorem

$$= LHS$$

$$E_Y[E_{X|Y}[f(X,Y)|Y]] = \int E_{X|y}[f(X,y)|y]\, g(y)\, dy$$

$$= \iint f(x,y)\, g(x|y)\, g(y)\, dy\, dx$$

$$= \iint f(x,y)\, g(x,y)\, dy\, dx$$

$$= LHS$$

Q2. N: Random Variable (RV) for no. of accidents.

X_i : RV for no. of injured in accident i.

We need to find $E[\sum_{i=1}^{N} X_i]$.

Since N itself is a random variable, we use total expectation rule. $E[\sum_{i=1}^{N} X_i] = E_N [E_{X|N}[\sum_{i=1}^{N} X_i | N]]$

$$E_{X|N=n}\left[\sum_{i=1}^{N=n} X_i \mid N=n\right]$$

$$= \sum_{i=1}^{n} E_{X|n}[X_i|n] = n\,E[X]$$

$\left(\because X_i\text{'s are identically distributed \& } X_i \text{ is independent of } N \right)$

$$\Rightarrow E[\sum_{i=1}^{N} X_i] = E_N[N\,E[X]]$$

$$= E[N]\,E[X]$$

$$= \boxed{8}$$

Q3. Marginal Gaussian:

$$p(x,y) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left((z-\mu_z)^T \Sigma^{-1} (z-\mu_z)\right)\right\}$$

where $z = \begin{bmatrix} x \\ y \end{bmatrix}$ & $\mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$

& $\Sigma^{-1} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}^{-1}$

$= \begin{bmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{bmatrix}$

$$p(x,y) \propto \exp\left\{-\frac{1}{2}\left((x-\mu_x)^T \Lambda_{xx} (x-\mu_x) + (y-\mu_y)^T \Lambda_{yy} (y-\mu_y) + 2(x-\mu_x)^T \Lambda_{xy} (y-\mu_y)\right)\right\} \quad \text{--- } ①$$

To show that the marginal $p(x)$ is Gaussian, we can just show

$$p(x) \propto \exp\left\{-\frac{1}{2}(x-t)^T A^{-1} (x-t)\right\}$$

where A is some PSD matrix.

$$p(x) = \int p(x,y)\, dy$$

To integrate out y, we will complete square in y because $\int e^{-(y-s)^T B^{-1} (y-s)} dy$ for some PSD matrix B is $(2\pi)^{m/2} |B|^{1/2}$ which doesn't contain x & hence can be ignored.

Collecting y terms in the quadratic in ①:

$$(y-\mu_y)^T \Lambda_{yy} (y-\mu_y) + 2(y-\mu_y)^T \Lambda_{yx} (x-\mu_x)$$

By inspection, we can write this as

$$\left(y-\mu_y + \Lambda_{yy}^{-1} \Lambda_{yx} (x-\mu_x)\right)^T \Lambda_{yy} \left(y-\mu_y + \Lambda_{yy}^{-1} \Lambda_{yx} (x-\mu_x)\right)$$

$$- (x-\mu_x)^T \Lambda_{xy} \Lambda_{yy}^{-1} \Lambda_{yx} (x-\mu_x)$$

Hence, after integrating out y,

$p(x) \propto$

$$\exp\left(-\frac{1}{2} (x-\mu_x)^T \left(\Lambda_{xx} - \Lambda_{xy} \Lambda_{yy}^{-1} \Lambda_{yx}\right)(x-\mu_x)\right)$$

By Schur Complement,

$$\Lambda_{xx} - \Lambda_{xy} \Lambda_{yy}^{-1} \Lambda_{yx} = \Sigma_{xx}^{-1}$$

$$\therefore p(x) \propto \exp\left(-\frac{1}{2} (x-\mu_x)^T \Sigma_{xx}^{-1} (x-\mu_x)\right)$$

Similary, $p(y) \propto \exp\left(-\frac{1}{2} (y-\mu_y)^T \Sigma_{yy}^{-1} (y-\mu_y)\right)$

Conditional Gaussian

To show $p(x|y)$ is Gaussian, we can just show

$$p(x|y) \propto \exp\left(-\frac{1}{2}(x-\mu(y))^T \Sigma(y)^{-1}(x-\mu(y))\right)$$

$p(x|y) \propto p(x,y)$

As $p(x|y)$ is a distribution over x, we ignore terms in the quadratic of ① that don't involve x & then complete square in x.

$$(x-\mu_x)^T \Lambda_{xx}(x-\mu_x) + 2(x-\mu_x)^T \Lambda_{xy}(y-\mu_y)$$

$$\propto (x-\mu_x+\Lambda_{xx}^{-1}\Lambda_{xy}(y-\mu_y))^T \Lambda_{xx} (x-\mu_x+\Lambda_{xx}^{-1}\Lambda_{xy}(y-\mu_y))$$

$\therefore p(x|y) \propto$

$$\exp\left\{-\frac{1}{2}(x-\mu_x+\Lambda_{xx}^{-1}\Lambda_{xy}(y-\mu_y))^T \Lambda_{xx} (x-\mu_x+\Lambda_{xx}^{-1}\Lambda_{xy}(y-\mu_y))\right\}$$

From Schur Complement,

$$\Lambda_{xx} = (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}$$

(From Schur Complement Lemma, Λ_{xx}^{-1} is PSD as Σ is PSD.)

$$\Lambda_{xy} = -\Lambda_{xx}\Sigma_{xy}\Sigma_{yy}^{-1}$$

$$\therefore X|y \sim N(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Lambda_{xx}^{-1})$$

Similarly we can prove for $Y|x$.

Q4 $Y \in \mathbb{R}^{m\times 1}$

$$E[Y]_i = E[Y_i]$$

$$= E[w_i^T X + b]$$

where w_i is the ith row of W

$$= w_i^T E[X] + b$$

$$E[Y] = W E[X] + b$$

$$Cov(Y) = E[(Y - E[Y])(Y - E[Y])^T]$$

$$= E[(WX - W E[X])(WX - WE[X])^T]$$

$$= E[W(X - E[X])(X - E[X])W^T]$$

$$= W\, Cov(X)\, W^T$$

1.2) Q1 (a)

$$f^*(x) = \underset{y\in\{0,1\}}{\operatorname{argmin}} \; E_{Y|x}[\ell(y,Y)|x]$$

$$= \underset{y\in\{0,1\}}{\operatorname{argmin}} \; P[y \neq Y|x]$$

$$= \underset{y\in\{0,1\}}{\operatorname{argmin}} \; 1 - P[Y=y|x]$$

$$= \underset{y\in\{0,1\}}{\operatorname{argmax}} \; P[Y=y|x]$$

$$\therefore \quad f^*(x) = \begin{cases} 1 & \forall x \text{ s.t. } p^*(1|x) \geq p^*(0|x) \\ 0, & \forall x \text{ s.t. } p^*(0|x) > p^*(1|x) \end{cases}$$

$$p^*(1|x) = \frac{1}{1+e^{x^2-5x+6}}$$

$p^*(1|x) \geq p^*(0|x)$ whenever

$$e^{x^2-5x+6} \leq e^0$$

$$\Rightarrow x^2 - 5x + 6 \leq 0$$

(as $\exp(\cdot)$ is a monotonically increasing function)

$$\therefore f^*(x) = \begin{cases} 1 \text{ when } x \in [2,3] \\ 0 \text{ when } x \geq 3 \text{ or } x \leq 2 \end{cases}$$

Q1 (b)

$$f^*(x) = \underset{y\in\{0,1\}}{\operatorname{argmin}} \; E_{Y|x}\left[\ell(y,Y)|x\right]$$

$$= \underset{y\in\{0,1\}}{\operatorname{argmin}} \; \ell(y,0)\, p^*(0|x) + \ell(y,1)\, p^*(1|x)$$

Let us call $\ell(y,0)p^*(0|x)+\ell(y,1)\,p^*(1|x)$ as $h_x(y)$

$$f^*(x) = \underset{y\in\{0,1\}}{\operatorname{argmin}} \; h_x(y)$$

$$h_x(0) = \ell(0,0)\, p^*(0|x) + \ell(0,1)\, p^*(1|x)$$
$$= \frac{0.5}{1+e^{x^2-5x+6}}$$

$$h_x(1) = \ell(1,0)\, p^*(0|x) + \ell(1,1)\, p^*(1|x)$$
$$= \frac{2\, e^{x^2-5x+6}}{1+e^{x^2-5x+6}}$$

It is easy to see that $0.5 < 2e^{x^2-5x+6}$ because for the quadratic $x^2-5x+6+\ln\left(\frac{2}{0.5}\right)$ coefficient of x^2 i.e. 1 is positive & the discriminant $25-4(6+\ln 4)$ is negative.

Hence, $h_x(0) < h_x(1) \quad \forall x \in \mathbb{R}$

$$\therefore f^*(x) = 0 \quad \forall x \in \mathbb{R}$$

Q1(c)

$$f^*(x) = \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; E_{Y|x}[\ell(y, Y) | x]$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; E_{Y|x}[(y - Y)^2 | x]$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; E_{Y|x}[y^2 + Y^2 - 2yY \,|x]$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; y^2 - 2y\, E_{Y|x}[Y|x] + E_{Y|x}[Y^2|x]$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; \left(y - E_{Y|x}[Y|x]\right)^2$$

(after completing square in y & removing terms not having y)

$$= E_{Y|x}[Y|x]$$

Given: $p^*(y|x) \propto e^{-\lambda(x) y}, \quad y > 0$

$$\Lambda(x) = \max\text{eig}\left(\sum_{i=1}^{n} A_i x_i + A_0\right)$$

Eigenvalues of $\sum_{i=1}^{n} A_i x_i + A_0 = \sum_{i=1}^{n} \lambda_i x_i + \lambda_0$
where λ_i is an eigenvalue of A_i &
λ_0 is an eigenvalue of A_0

By definition of PSD matrices, λ_i's are non-negative & by definition of PD matrices, λ_0 is positive.

Hence $\Lambda(x) = \max\text{eig}\left(\sum_{i=1}^{n} A_i x_i + A_0\right) > 0$

$$p^*(y|x) = \frac{e^{-\Lambda(x)y}}{\int e^{-\Lambda(x)y}} = \Lambda(x)\, e^{-\Lambda(x)y}$$

this is P.D.F. of an exponential distribution.

$$\therefore f^*(x) = E_{Y|x}[Y|x]$$

$$= \int_0^\infty \Lambda(x)\, y e^{-\Lambda(x)y}\, dy$$

$$= \Lambda(x)\left(\left|-\frac{y\, e^{-\Lambda(x)y}}{\Lambda(x)}\right|_0^\infty + \int_0^\infty \frac{e^{-\Lambda(x)y}}{\Lambda(x)}\, dy\right)$$

$$= 1/\Lambda(x)$$

Q1 (d)

$$f^*(x) = \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \; E_{Y|x} [\, l(y, Y) | x \,]$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \int_0^\infty l(y, y')\, p(y'|x)\, dy'$$

$$= \underset{y \in \mathbb{R}_{++}}{\operatorname{argmin}} \int_0^\infty |y - y'|\, p(y'|x)\, dy'$$

(Extra detail :
As the objective is convex in y & the domain is open, the argmin y should satisfy sub-gradient wrt y as 0 condition)

Let us call the objective $\int_0^\infty |y-y'|\, p(y'|x) dy'$ as $h_x(y)$

$$h_x(y) = \left(\int_0^y (y - y')\, p(y'|x)\, dy' + \int_y^\infty (y' - y)\, p(y'|x)\, dy' \right)$$

we will use Leibnitz integral rule for differentiation under the integral sign to differentiate $h_x(y)$ wrt y.

$$\frac{d}{dy} h_x(y) = \int_0^y 1 \cdot p(y'|x)\, dy' + \int_y^\infty (-1)\, p(y'|x)\, dy'$$

Let the optimal (argmin) y be denoted as y^* then

$$\int_0^{y^*} p(y'|x)\, dy' = \int_{y^*}^\infty p(y'|x)\, dy'$$

$$\Rightarrow F_{Y|x}(y^*) = 1 - F_{Y|x}(y^*)$$

$$\Rightarrow F_{Y|x}(y^*) = 0.5$$

$$\Rightarrow \int_0^{y^*} \lambda(x)\, e^{-\lambda(x) y^*}\, dy = 0.5$$

$$\Rightarrow \left. -e^{-\lambda(x) y^*} \right|_0^{y^*} = 0.5$$

$$\Rightarrow e^{-\lambda(x) y^*} = 0.5$$

$$\Rightarrow y^* = \frac{\ln 2}{\lambda(x)}$$

$$\therefore f^*(x) = \ln 2 / \lambda(x)$$