«به نام خدا»

یادگیری ماشین

امیررضا آذری

تمرین نظری اول

۹۹۱۰۱۰۷

① رگرسیون خطی:

دو حالت $N \leq M$ و $M \leq N$ را در نظر می‌گیریم. داریم:

① $N > M$:

خطی بودن E

$$E[R_{tr}(\hat{\beta})] = E\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}^T x_i)^2\right] = \frac{1}{N}\sum_{i=1}^{N} E(y_i - \hat{\beta}^T x_i)^2$$

i.i.d

$$= E(y_1 - \hat{\beta}^T x_1)^2 = \frac{1}{M}\sum_{i=1}^{M} E(y_i - \hat{\beta}^T x_i)^2$$

حال اگر $\beta' = \underset{\beta}{\operatorname{argmin}} \frac{1}{M}\sum_{i=1}^{M} E(y_i - \beta^T x_i)^2$ در نظر بگیریم، داریم:

$$\frac{1}{M}\sum_{i=1}^{M} E(y_i - \hat{\beta}^T x_i)^2 \leq \frac{1}{M}\sum_{i=1}^{M} E(y_i - \hat{\beta}'^T x_i)^2 = \frac{1}{M}\sum_{i=1}^{M} E(\tilde{y}_i - \tilde{\beta}^T \tilde{x}_i)^2$$

$$\leq \frac{1}{M}\sum_{i=1}^{M} E(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2$$

$$= E\left(\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2\right) = E[R_{te}(\hat{\beta})]$$

در نوشتن معادلات و نامعادلات از i.i.d بودن و خطی بودن امید ریاضی بهره بردیم.

بنابراین برای $N > M$ داریم:

$$E[R_{tr}(\hat{\beta})] \leq E[R_{te}(\hat{\beta})]$$

۲ $N \le M$:

$$E[R_{te}(\hat\beta)] = E\left(\frac{1}{M}\sum_{i=1}^{M}(\tilde y_i - \hat\beta^T \tilde x_i)^2\right) = \frac{1}{M}\sum_{i=1}^{M} E(\tilde y_i - \hat\beta^T \tilde x_i)^2$$

$$\ge \frac{1}{M}\sum_{i=1}^{M} E(\tilde y_i - \tilde\beta^T \tilde x_i)^2$$

که $\tilde\beta = \underset{\beta}{\arg\min} \frac{1}{M}\sum_{i=1}^{M} E(\tilde y_i - \beta^T \tilde x_i)^2$ ، ادامه می دهیم :

$$\frac{1}{M}\sum_{i=1}^{M} E(\tilde y_i - \tilde\beta^T \tilde x_i)^2 = E(\tilde y_1 - \tilde\beta^T \tilde x_i)^2 = \frac{1}{N}\sum_{i=1}^{N} E(\tilde y_i - \tilde\beta^T \tilde x_i)^2$$

$$\ge \frac{1}{N}\sum_{i=1}^{N} E(\tilde y_i - \beta'^T \tilde x_i)^2$$

که $\beta' = \underset{\beta}{\arg\min} \frac{1}{N}\sum_{i=1}^{N} E(\tilde y_i - \beta^T \tilde x_i)^2$ ، ادامه می دهیم :

$$\frac{1}{N}\sum_{i=1}^{N} E(\tilde y_i - \beta'^T \tilde x_i)^2 = \frac{1}{N}\sum_{i=1}^{N} E(y_i - \hat\beta^T x_i)^2 = E\left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat\beta^T x_i)^2\right)$$

$$= E[R_{tr}(\hat\beta)]$$

در تمامی نوشتن مطالب و معادلات از فرض iid بودن $(\tilde x_i, \tilde y_i)$ و (x_i, y_i) استفاده کردیم.

در بخش اول هم می‌دانیم $\hat\beta = \underset{\beta}{\arg\min} \frac{1}{N}\sum_{i=1}^{N} E(y_i - \beta^T x_i)^2$

بنابراین در هر دو حالت اثبات کردیم :

$$E[R_{tr}(\hat\beta)] \le E[R_{te}(\hat\beta)]$$

② رگرسیون خطی:

آ) می دانیم:

$$w^* = \arg\min J_\lambda(w)$$

$$J_\lambda = \frac{1}{2}\|y - Xw\|_2^2 + \lambda \|w\|_1$$

$$X^T X = I$$

حال J_λ را باز می کنیم:

$$J_\lambda = \frac{1}{2}(y^T y - 2y^T X w + w^T X^T X w) + \lambda \|w\|_1$$

با $X^T X = I$:

$$J_\lambda = \frac{1}{2}(y^T y - 2y^T X w + w^T w) + \lambda \|w\|_1$$

$$\rightarrow \lambda \|w\|_1 = \sum_{i=1}^{d} \lambda |w_i|$$

$$\rightarrow w^T w = \|w\|^2 = \sum_{i=1}^{d} w_i^2$$

$$\rightarrow -2y^T X w = \sum_{i=1}^{d} -2y^T X_{:,i} w_i$$

$$\Rightarrow J_\lambda = \underbrace{\frac{1}{2} y^T y}_{g(y)} + \underbrace{\sum_{i=1}^{d} -y^T X_{:,i} w_i + \frac{1}{2} w_i^2 + \lambda |w_i|}_{\sum_{i=1}^{d} f(X_{:,i}, y, w_i, \lambda)}$$

ب) اگر $w_i > 0$، به شکل بخش قبل عبارت $\ell(x_{:,i}; y; w_i; \lambda)$ را minimize کنیم:

$$\frac{\partial \ell(x_{:,i}; y; w_i; \lambda)}{\partial w_i} = 0 \rightarrow w_i^* - y^T x_{:,i} + \lambda = 0 \Rightarrow w_i^* = y^T x_{:,i} - \lambda \quad (y^T x_{:,i} > \lambda)$$

پ) اگر $w_i < 0$، مانند بخش قبل داریم:

$$\frac{\partial \ell}{\partial w_i} = 0 \xRightarrow{|w_i| = -w_i} \frac{\partial}{\partial w_i}\left(\frac{1}{2} w_i^2 - y^T x_{:,i} w_i - \lambda w_i\right) = 0 \Rightarrow w_i^* - y^T x_{:,i} - \lambda = 0$$

$$\Rightarrow w_i^* = y^T x_{:,i} + \lambda \quad (y^T x_{:,i} < -\lambda)$$

ت)

با توجه به قسمت های قبل، برای صفر شدن w_i داریم:

$$y^T x_{:,i} - \lambda \le 0 \quad \text{و} \quad y^T x_{:,i} + \lambda \ge 0$$

که نتیجه می شود:

$$-\lambda \le y^T x_{:,i} \le \lambda$$

ث)

در این حالت داریم:

$$\ell(x_{:,i}; y; w_i; \lambda) = \frac{1}{2} w_i^2 - y^T x_{:,i} w_i + \frac{\lambda}{2} w_i^2$$

$$\rightarrow \frac{\partial \ell}{\partial w_i} = 0 \rightarrow w_i - y^T x_{:,i} + \lambda w_i = 0$$

$$\rightarrow w_i = \frac{y^T x_{:,i}}{\lambda + 1}$$

(و یا $\lambda \to \infty$)

که در حالتی که $y^T x_{:,i}$ صفر شود، w_i صفر می شود. در حالت لاسو w_i زمانی صفر می شود که $-\lambda \le y^T x_{:,i} \le \lambda$ باشد. به همین دلیل است که regularization L_1-norm دارای sparsity است.

③ دسته بندی خطی:

داریم:

$w^0 = 0$

$w^{t+1} = w^t + \eta x^t y^t$

$$\Rightarrow w^1 = w^0 + \eta x^0 y^0 = \eta x^0 y^0$$
$$w^2 = w^1 + \eta x^1 y^1 = \eta x^0 y^0 + \eta x^1 y^1$$
$$w^3 = w^2 + \eta x^2 y^2 = \eta x^0 y^0 + \eta x^1 y^1 + \eta x^2 y^2$$
$$\vdots$$
$$w^t = \eta x^0 y^0 + \eta x^1 y^1 + \dots + \eta x^{t-1} y^{t-1} \Rightarrow w^t = \sum_{i=0}^{t-1} (l_i)\, \eta x^i y^i$$

که l_i دفعاتی است که سمپل iام را آپدیت کرده‌ایم. داریم:

$$w = \sum_{i=1}^{N} (l_i)\, \eta x^i y^i = w = \sum_{i=1}^{N} \alpha_i x^i$$

که $\alpha_i = l_i \eta y^i$ و اثبات کردیم که w را به صورت ترکیب خطی داده‌های x^i می‌توان نوشت.

④

الف)

$$P(y=1 \mid x) = \frac{P(x \mid y=1) \times P(y=1)}{P(x)} = \frac{P(c_1, c_2, \dots, c_d \mid y=1) \times P(y=1)}{P(c_1, \dots, c_d)}$$

$$= P(y=1) \times \frac{\prod_{i=1}^{d} P(c_i \mid y=1)}{\prod_{i=1}^{d} P(c_i)} \qquad (I)$$

طبق صورت سوال می‌دانیم:

* $P_y = P(y=1)$

* $P_{i|y=1} = P(c_i \mid y=1)$

بنابراین :

$$(\Sigma) = P_y \times \frac{\prod_{i=1}^{d} P_{i|y=1}}{\prod_{i=1}^{d} p(c_i)}$$

بنابراین $p(y=1|X)$ را به حالت بالا درآوردیم که با $P_y \prod_{i=1}^{d} P_{i|y=1}$ متناسب است. (راه دیگر ۴ الف بعد از ۴ ج نوشته شده است)

ب)

اکنون باید نشان بدهیم که مرز این مدل طبقه‌بند خطی است.

مرز تصمیم $\rightarrow$ $p(y=0|x) = p(y=1|x)$

$$\Rightarrow p(y=0|x) + p(y=1|x) = 1 \longrightarrow p(y=1|x) = \frac{1}{2}$$

حال ادامه می‌دهیم :

$$p(y=1|x) = \frac{1}{2} = \frac{1}{1+T}$$

که در آن T برابر زیر است :

$$T = \frac{1-P_y}{P_y}\left(\prod_{i=1}^{d}\left(\frac{P_{i|y=0}}{P_{i|y=1}}\right)^{c_i}\right) = 1$$

می‌دانیم log یک monotonic است : داریم :

$$\Rightarrow \underbrace{\log\left(\frac{1-P_y}{P_y}\right)}_{a_2} + \sum_{i=1}^{d} c_i \underbrace{\log\left(\frac{P_{i|y=0}}{P_{i|y=1}}\right)}_{w_i} = 0 \implies a_2 + w^T x = 0$$

در نتیجه مرز تصمیم‌گیری خطی است!

s.a.m

ج)

یک بار سوال الف را با یک راه حل کردیم، بار دیگر سوال الف را بعد از این بخش حل کرده ایم و اکنون از آن کمک می‌گیریم:

می‌خواهیم نشان بدهیم:

$$P(y=1\mid x)=\frac{1}{1+e^{-(\theta^T x+\theta_0)}}=\frac{1}{1+e^{-(\theta_0+\sum\theta_i x_i)}}$$

طبق عبارت بخش الف ←

$$\Rightarrow \frac{1}{1+e^{t}}\ ,\ \text{where } t=\ln\left(\frac{1-P_y}{P_y}\left(\Pi\left(\frac{P_{i\mid y=0}}{P_{i\mid y=1}}\right)^{c_i}\right)\right)$$

$$=\frac{1}{1+e^{t'}}\ ,\ \text{where } t'=-\ln\left(\frac{P_y}{1-P_y}\right)-\sum c_i\ln\left(\frac{P_{i\mid y=1}}{P_{i\mid y=0}}\right)$$

همانطور که در t' مشاهده می‌کنید، داریم:

$$\theta_0=\ln\left(\frac{P_y}{1-P_y}\right)\quad \text{و}\quad \theta^T x=\sum_{i=1}^{d}c_i\ln\left(\frac{P_{i\mid y=1}}{P_{i\mid y=0}}\right)\Rightarrow\frac{1}{1+e^{-(\theta_0+\theta^T x)}}$$

بنابراین احتمال شرطی بخش الف، تابع logistic می‌باشد.

$\boxed{\text{۲. الف}}$

قضیه بیز و احتمال کل را که می‌دانیم، اگر آن را به عبارت حاضر در صورت تقسیم کنیم داریم:

$$P(y=1\mid x)=\frac{1}{1+\frac{P(y=0)\,P(x\mid y=0)}{P(y=1)\,P(x\mid y=1)}}$$

به دلیل استقلال توضیح داده شده در بخش ۴ الف ابتدا داریم:

$$p(x|y=0) = p_{1|y=0}\, p_{2|y=0} \cdots = \prod_{i=1}^{d} p_{i|y=0}^{c_i}$$

$$p(y=1|x) = \frac{1}{1 + \frac{1-p_y}{p_y}\left(\prod_{i=1}^{d}\left(\frac{p_{i|y=0}}{p_{i|y=1}}\right)^{c_i}\right)}$$

(ه)

الف)

می دانیم $P(w_1)=P(w_2)=P(w_3)=\frac{1}{3}$. داریم:

$$P(w_i|x)=\frac{P(x|w_i)P(w_i)}{P(x)}.$$

$$\rightarrow P(x|w_1)=\frac{1}{2\pi}\exp\left[-\frac{1}{2}(0.3, 0.3)[0.3, 0.3]^T\right]=0.1455$$

$$\rightarrow P(x|w_2)=\frac{1}{2\pi}\exp\left[-\frac{1}{2}[-0.7 \quad -0.7][-0.7 \quad -0.7]^T\right]=0.0975$$

$$\rightarrow P(x|w_3)=\frac{1}{4\pi}\exp\left[-\frac{1}{2}[-0.2 \quad -0.2][-0.2 \quad -0.2]^T\right]+\frac{1}{4\pi}\exp\left[-\frac{1}{2}[0.8 \quad -0.2][0.8 \quad -0.2]^T\right]$$

$$=0.1331$$

از آنجا که $P(x|w_1)$ بیشترین مقدار likelihood را دارد، x متعلق به کلاس ۱ خواهد بود.

محاسبات بیشتر:

$$\rightarrow P(w_1|x)=\frac{\frac{1}{3}\times 0.1455}{\frac{1}{3}(0.1455+0.0975+0.1331)}=\frac{\frac{1}{3}\times 0.1455}{0.1253}\simeq 0.38707$$

$$\rightarrow P(w_2|x)=\frac{\frac{1}{3}\times 0.0975}{0.1253}\simeq 0.259377$$

$$\rightarrow P(w_3|x)=\frac{\frac{1}{3}\times 0.1331}{0.1253}\simeq 0.354083$$

پس X متعلق به دسته کلاس ۱ است. (در این صورت کمترین احتمال خطا را داریم)

ب)

$$x = [x_1, x_2]^t \rightarrow p(w_i \mid x_2) = \frac{\int p(x_1, x_2 \mid w_i)\, p(w_i)\, dx_1}{p(x_2)}$$

$$\longrightarrow \int p(x_1, x_2 \mid w_1)\, dx_1 = \frac{1}{2\pi} \int \exp\left[-\frac{1}{2}(x_1^2 + 0.09)\right] dx_1$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{0.09}{2}\right) = 0.3813$$

$$\longrightarrow \int p(x_1, x_2 \mid w_2)\, dx_1 = \frac{1}{2\pi} \int \exp\left[-\frac{1}{2}\left((x_1 - 1)^2 + 0.49\right)\right] dx_1$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-0.49}{2}\right) = 0.312$$

$$\longrightarrow \int p(x_1, x_2 \mid w_3)\, dx_1 = \frac{1}{4\pi} \int \exp\left[-\frac{1}{2}\left((x_1 - 0.5)^2 + 0.04\right)\right] dx_1$$

$$+ \frac{1}{4\pi} \int \exp\left[-\frac{1}{2}\left((x_1 + 0.5)^2 + 0.04\right)\right] dx_1$$

$$= 0.391$$

بنابراین این نقطه به کلاس ۳ تعلق دارد.

۶

الف) این سؤال را مطابق کلاس حل تمرین حل می‌کنیم.

W را باید طوری بدست بیاوریم که:

$$\frac{1}{2} \sum_{i=1}^{m} w(i) \left(\theta^T x^{(i)} - y^{(i)}\right)^2 = (X\theta - y)^T W (X\theta - y)$$

$$\Rightarrow \sum_{i=1}^{m} \left(\theta^T x^{(i)} - y^{(i)}\right)^2 = (X\theta - y)^T (X\theta - y)$$

نتیجه می‌گیریم که W به دلیل وجود ۲ عامل $w(i)$ و $\frac{1}{2}$ بدست آمده است. در نتیجه اعمالش هم از $\frac{1}{2}$ و $w(i)$

s.a.m

می‌آیند. زیرا اگر این ها دیر نبودند، حاصل $\sum(\theta^T x^{(i)} - y^{(i)})^2$ برابر $(X\theta - y)^T(X\theta - y)$ می‌شد.

از طرف دیگر می‌دانیم که داده‌ها تحت درهم ضرب نمی‌شوند و داده‌ها از هم مستقل هستند، بنابراین W قطری است.

در نتیجه‌گیری حرف پایانی خود داریم:

$$W = \begin{bmatrix} \frac{1}{2}w^{(i)} & 0 & \cdots & 0 \\ 0 & \frac{1}{2}w^{(i)} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \frac{1}{2}w^{(i)} \end{bmatrix}$$

پس ماتریس W را بدست آوردیم.

$$W = \begin{cases} \frac{1}{2}w^{(i)} & i=j \\ 0 & \text{o.w.} \end{cases}$$

ب)

برای این بخش داریم:

$$\nabla_\theta J(\theta) = \frac{\partial}{\partial\theta}\left((X\theta - y)^T W (X\theta - y)\right) = \frac{\partial}{\partial\theta}\left((\theta^T X^T - y^T) W (X\theta - y)\right)$$

$$= \frac{\partial}{\partial\theta}\left((\theta^T X^T - y^T)(WX\theta - Wy)\right) = \frac{\partial}{\partial\theta}\left(\theta^T X^T W X\theta - \theta^T X^T W y - y^T W X\theta + y^T W y\right)$$

* نکته ۱: می‌دانیم W متقارن است. $(W^T = W)$.
* نکته ۲: ″ ″ $(\theta^T X^T W y)^T = y^T W X\theta$.

بنابراین داریم:

$$\nabla_\theta J(\theta) = \frac{\partial}{\partial\theta}\left(\theta^T X^T W X\theta - 2y^T W X\theta + y^T W y\right)$$

* نکته ۳: $\nabla_\theta(\theta^T A\theta) = (A + A^T)\theta$
* نکته ۴: $\nabla_\theta(A\theta) = A^T\theta$

بنابراین داریم:

$$\nabla_\theta J(\theta) = (X^T W X + X^T W X)\theta - 2X^T W y = 2X^T W X\theta - 2X^T W y = 0$$

$$\rightarrow X^T W X\theta = X^T W y$$

$$\Rightarrow \theta = (X^T W X)^{-1} X^T W y$$

که اگر همه وزن‌ها یکی باشند، $W = I$ و به نکته گفته شده در سوال می‌رسیم.

s.a.m

ج)

$$L(\theta) = \ln\left(\prod_{i=1}^{m} \left(p(y^{(i)} \mid x^{(i)}; \theta)\right)\right) = \sum_{i=1}^{m} \ln\left(p(y^{(i)} \mid x^{(i)}; \theta)\right)$$

$$= \sum_{i=1}^{m} \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)\right)$$

$$= \sum_{i=1}^{m} \left(-\ln(\sqrt{2\pi}\,\sigma^{(i)}) - \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

$$\longrightarrow \frac{\partial L(\theta)}{\partial \theta} = -\sum_{i=1}^{m} \frac{\partial}{\partial \theta}\left(\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

با سادگی نسبت به درایه‌ی j ام θ مشتق می‌گیریم:

$$\longrightarrow \frac{\partial L(\theta)}{\partial \theta_j} = -\sum_{i=1}^{m} \frac{(y^{(i)} - \theta^T x^{(i)})}{(\sigma^{(i)})^2} x_j^{(i)}$$

حال عبارت $L(\theta)$ را با عبارت داخل صورت سؤال مقایسه می‌نماییم. در نتیجه:

$$w^{(i)} = -\frac{1}{(\sigma^{(i)})^2}$$

در واقع داریم:

$$\frac{\partial J(\theta)}{\partial \theta} = \sum_{i=1}^{m} \frac{w(i)}{2} \times 2\left[\theta^T x^{(i)} - y^{(i)}\right] x^{(i)} = -\sum_{i=1}^{m} \frac{(y^{(i)} - \theta^T x^{(i)})}{(\sigma^{(i)})^2} x^{(i)}$$

$$\Rightarrow w(i) = -\frac{1}{(\sigma^{(i)})^2}$$

۷

الف)

$$P_{X_1|Y}(x_1|y=c)=\begin{cases}\theta_c & x_1=1\\ 1-\theta_c & \text{o.w.}\end{cases}$$

$$\rightarrow P_{X_1|Y}(x_1|y=c)=\theta_c x_1+(1-x_1)(1-\theta_c)$$

$$\rightarrow P_{X_2|Y}(x_2|y=c)=\frac{1}{\sqrt{2\pi}\,\sigma_c}e^{-\frac{(x_2-\mu_c)^2}{2\sigma_c^2}}$$

طبق بیز داریم:

$$p(\theta|D)=\frac{p(D|\theta)\,p(\theta)}{p(D)}$$

در نتیجه:

$$P_{Y|X_1,X_2}(y|x_1=0,x_2=0)=\frac{P_Y(y)\,p(x_1=0,x_2=0|y)}{p(x_1=0,x_2=0)}$$

Naive Bayes $\rightarrow$ $P_{Y|X_1,X_2}(y|x_1=0,x_2=0)\propto P_Y(y=0)\,p(x_1=0|y=0)\,p(x_2=0|y=0)$

$*\ P_Y(y=0)=0.5$ $\quad *\ p(x_2=0|y=0)=\text{norm}(0;-1,1)=0.242=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2})$

$*\ p(x_2=0|y=1)=\frac{1}{\sqrt{2\pi}}\exp(0)=0.4$

$*\ p(x_2=0|y=2)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2})=0.242$

$$\rightarrow P_{Y|X_1,X_2}(y=0|x_1=0,x_2=0)=0.0604$$

$$P_{Y|X_1,X_2}(y=1|x_1=0,x_2=0)=0.0498$$

$$P_{Y|X_1,X_2}(y=2|x_1=0,x_2=0)=0.0302$$

حال باید نرمالایز کنیم:

$$\sum_{n=0}^{2} p(x_1=0|y=n)\,p(x_2=0|y=n)\,p(c)=\frac{1}{2}\left(0.242\times0.15+0.4\times0.25+0.242\times0.25\right)$$

$$=0.1407$$

$$\Longrightarrow P_{Y|X_1,X_2}(y|x_1=0,x_2=0)=\frac{1}{0.1407}\begin{bmatrix}0.0604\\0.0498\\0.0302\end{bmatrix}=\begin{bmatrix}0.43\\0.36\\0.21\end{bmatrix}=\begin{bmatrix}0.43\\0.355\\0.215\end{bmatrix}$$

ب)

داریم:

$$P_{Y|X_1}(y|x_1=0)\propto P_{X_1|Y}(x_1=0|y)\,p(y)$$

$$\longrightarrow P_{Y|X_1}(x_1=0|y=0)\,p(y=0)=1/4$$

$$P_{Y|X_1}(x_1=0|y=1)\,p(y=1)=1/8$$

$$P_{Y|X_1}(x_1=0|y=2)\,p(y=2)=1/8$$

حال نرمالایز می‌کنیم:

$$\sum_{n=0}^{2} p(x_1=0|y=n)\,p(y=n)=\frac{1}{2}$$

$$\Rightarrow P_{Y|X_1}(y|x_1=0)=2\begin{bmatrix}1/4\\1/8\\1/8\end{bmatrix}=\begin{bmatrix}1/2\\1/4\\1/4\end{bmatrix}$$

ب)

همانند بخش قبل:

$$\to p(x_2=0|y=0)\,p(y=0) = \frac{1}{2}\times 0.242 = 0.121$$

$$p(x_2=0|y=1)\,p(y=1) = \frac{1}{4}\times 0.4 = 0.1$$

$$p(x_2=0|y=2)\,p(y=2) = \frac{1}{4}\times 0.242 = 0.0605$$

باز هم نرمالایز کل اینجا می‌دهیم:

$$\sum_{n=0}^{2} p(x_2=0|y=n)\,p(y=n) = 0.2815$$

$$\Longrightarrow P_{Y|X_2}(y|x_2=0) = \frac{1}{0.2815}\begin{bmatrix} 0.121 \\ 0.1 \\ 0.0605 \end{bmatrix} = \begin{bmatrix} 0.4298 \\ 0.3552 \\ 0.2149 \end{bmatrix}$$

ت)

مشاهده می‌کنیم که $P_{Y|X_1,X_2}(y|x_1=0, x_2=0) = P_{Y|X_2}(y|x_2=0)$.

بنابراین ویژگی $X_1=0$ زیرمجموعه ویژگی $X_2=0$ است و توزیع توأم آن‌ها همان $x_2=0$ شده است.