تمرین سری اول یادگیری ماشین

امیرمحمد عزتی 402212269

①

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

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

اگر $N \leqslant M$ باشد:

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

$$\geqslant \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}\left(\bar{y}_i - \bar{\beta}^T \bar{x}_i\right)^2$$

$\bar{\beta}$ پارامتری است که expected را روی داده‌های تست کمینه می‌کند.

$$= \mathbb{E}\left(\bar{y}_1 - \bar{\beta}^T \bar{x}_1\right)^2$$

اگر داده‌ها iid باشند

$$= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\bar{y}_i - \bar{\beta}^T \bar{x}_i\right)^2$$

$$\geqslant \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\bar{y}_i - \beta'^T \bar{x}_i\right)^2$$

β' پارامتری است که expected را روی N تا داده تست کمینه می‌کند

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

expected روی داده‌های train نیز همان مقدار است اگر داده‌های ترین و تست طبق گفته صورت سوال iid باشند.

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

$$\Rightarrow \mathbb{E}\left[R_{te}(\hat{\beta})\right] \geqslant \mathbb{E}\left[R_{tr}(\hat{\beta})\right]$$

اگر $N > M$ باشد:

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

$$= \mathbb{E}\left(y_1 - \hat{\beta}^T x_1\right)^2$$ → اگر داده‌ها iid باشند

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

$$\leq \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}\left(y_i - \beta'^T x_i\right)^2$$

β' پارامتری که روی expected را روی داده‌های ست کمینه می‌کند.

$$= \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}\left(\bar{y}_i - \bar{\beta}^T \bar{x}_i\right)^2$$ → iid بودن داده‌های train نسبت به test

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

$$= \mathbb{E}\left[R_{te}(\hat{\beta})\right]$$

$$\Rightarrow \quad \mathbb{E}\left[R_{tr}(\hat{\beta})\right] \leq \mathbb{E}\left[R_{te}(\hat{\beta})\right]$$

② همفکری با مرتضی شهرابی

الف)

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

$$J_\lambda(w) = \frac{1}{2} \sum_{j=1}^{d} (y_j - X_{:,j} w)^2 + \lambda \sum_{j=1}^{d} |w_j|$$

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

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

$$J_\lambda(w) = \frac{1}{2} \|y\|_2^2 - w^T X^T y + \frac{1}{2} w^T w + \lambda \|w\|_1$$

$$J_\lambda(w) = g(y) + \sum_{j=1}^{d} -w_j^T X_{j,:}^T y + \frac{1}{2} w_j^T w_j + \lambda |w_j|$$

$$J_\lambda(w) = g(y) + \sum_{j=1}^{d} f(X_{:,j} \,;\, y \,;\, w_j \,;\, \lambda)$$

در تابع f، مشخص است بهینه‌سازی w_j^* تحت تاثیر ویژگی j ام مدل و خروجی داده وابسته است.

ب) اگر $w_i \geq 0$ باشد آنگاه w_i

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

$$\xRightarrow{w_i \geq 0} \quad -X_{i,:}^T y + w_i + \lambda = 0 \Rightarrow w_i = X_{i,:}^T y - \lambda$$

پ) اگر $w_i < 0$ باشد، آنگاه w_i

$$\frac{\partial}{\partial w_i}\left(-w_i^T X_{i,:}^T y + \frac{1}{2} w_i^T w_i + \lambda |w_i|\right) = 0$$

$$\xRightarrow{w_i<0} \quad -X_{i,:}^T y + w_i - \lambda = 0 \Rightarrow w_i = X_{i,:}^T y + \lambda$$

ت)

$$w_i = X_{i,:}^T y - \lambda = 0 \Rightarrow X_{i,:}^T y = \lambda$$

زمانی w_i صفر خواهد شد که تاثیر ویژگی i ام برای تشخیص y ناچیز باشد.
یا ممکن است ویژگی i ام دارای Correlation بالای با بقیه فیچرها باشد
و برای ساده سازی و جلوگیری از over fitting مدل، از آن صرف نظر
شده است.

ث)

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

$$J_\lambda(w) = \frac{1}{2}\|y\|_2^2 - w^T X^T y + \frac{1}{2} w^T w + \frac{1}{2}\lambda w^T w$$

$$J_\lambda(w) = g(y) + \sum_{j=1}^{d} -w_j^T X_{j,:}^T y + \frac{1}{2} w_j^T w_j + \frac{1}{2}\lambda w_j^T w_j$$

$$\frac{\partial}{\partial w_i}\left(-w_j^T X_{j,:}^T y + \frac{1}{2} w_j^T w_j + \frac{1}{2}\lambda w_j^T w_j\right) = 0$$

$$\Rightarrow \quad -X_{j,:}^T y + w_j + \lambda w_j = 0 \Rightarrow w_j = \frac{X_{j,:}^T y}{1+\lambda}$$

زمانی که $X_{j,:}^T y = 0$ شود، آنگاه $w_j = 0$ خواهد بود.

③ اگر بروزرسانی وزن پرسپترون که بشکل زیر است را در نظر بگیریم :

$x^{(i)}$ misclassified $\qquad w^{(t+1)} = w^{(t)} + \eta\, x^{(i)} y^{(i)}$

اگر $w^{(0)} = 0$ در نظر بگیریم :

$$w^{(t+1)} = \eta \sum_{i=1}^{N} K_i\, x^{(i)} y^{(i)} \qquad \text{(به ازای N تا سمپل)}$$

✻ K_i تعداد باری است که سمپل $x^{(i)}$ اشتباه طبقه بندی شده است .

بطور مثال اگر در بار اول، سمپل $x^{(2)}$ اشتباه کلاس بندی شود

$$w^{(1)} = w^{(0)} + x^{(2)} y^{(2)} = \begin{cases} w^{(1)} = w^{(0)} - x^{(2)} & \text{if } y^{(2)} = -1 \\ w^{(1)} = w^{(0)} + x^{(2)} & \text{if } y^{(2)} = +1 \end{cases}$$

اگر بخواهیم بشکل دیگر نمایش دهیم :

$$w = \eta \sum_{i=1}^{N} K_i\, x^{(i)} y^{(i)} = \eta \left[\sum_{y^{(i)}=1} K_i x^{(i)} - \sum_{y^{(i)}=-1} K_i\, x^{(i)} \right]$$

پس اگر $w = \sum_{i=1}^{N} \alpha_i\, x^{(i)}$ باشد ، آنگاه ضرایب α_i بشکل زیر خواهد بود:

$$\alpha_i = \eta K_i\, y^{(i)}$$

K_i : تعداد بار اشتباه کلاس بندی شدن داده iام

η : learning rate

$y^{(i)}$: کلاس درست داده iام

④ الف)

$$P(y=1|x) = \frac{P(x|y=1)\ P(y=1)}{P(x)}$$

naive bayes

$$P(y=1|x) = \frac{\prod_{i=1}^{d} P(w_i|y=1)^{c_i} \cdot P(y=1)}{P(x)}$$

$P(w_i|y=1)$

احتمال حضور word i

در یک داکیومنت به شرط کلاس 1

ب) مرز تصمیم گیری مطابق روبه رو تعریف می شود:

$$P(y=0|x) = P(y=1|x) \Rightarrow \frac{P(y=1|x)}{P(y=0|x)} = 1$$

$$\frac{\prod_{i=1}^{d} P(w_i|y=1)^{c_i} \cdot P(y=1)}{\prod_{i=1}^{d} P(w_i|y=0)^{c_i} \cdot P(y=0)} = 1 \longrightarrow$$

$$\prod_{i=1}^{d} \left(\frac{P(w_i|y=1)}{P(w_i|y=0)}\right)^{c_i} \cdot \frac{P(y=1)}{P(y=0)} = 1 \xrightarrow{\log}$$

$$\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)} + \log \frac{P(y=1)}{P(y=0)} = 0$$

اگر c_i را همان ورودی x_i در نظر بگیریم و $w_i = \log \frac{P(w_i|y=1)}{P(w_i|y=0)}$ باشد

و همچنین مقدار ثابت $b = \log \frac{P(y=1)}{P(y=0)}$ باشد آنگاه:

$$\sum_{i=1}^{d} x_i w_i + b = 0 \Rightarrow \boxed{w^T x + b = 0}$$

پس مرز تصمیم خطی خواهد بود.

④ ج) از بخش قبل رابطه زیر را برای naive bayes داریم:

$$\log \frac{P(y=1|x)}{P(y=0|x)} = \sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)} + \log \frac{P(y=1)}{P(y=0)}$$

$$* \quad P(y=0|x) = 1 - P(y=1|x) \quad , \quad P(y=0) = 1 - P(y=1)$$

$$\log \frac{P(y=1|x)}{1-P(y=1|x)} = \sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)} + \log \frac{P(y=1)}{1-P(y=1)}$$

$$\xrightarrow{\exp} \frac{P(y=1|x)}{1-P(y=1|x)} = e^{\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)}} \times \frac{P(y=1)}{1-P(y=1)}$$

$$P(y=1|x) = \frac{e^{\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)}} \times \frac{P(y=1)}{1-P(y=1)}}{1 + e^{\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)}} \times \frac{P(y=1)}{1-P(y=1)}}$$

$$\Rightarrow P(y=1|x) = \frac{1}{\frac{1-P(y=1)}{P(y=1)} \times e^{-\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)}} + 1}$$

$$\Rightarrow P(y=1|x) = \frac{1}{1 + e^{-\log \frac{P(y=1)}{1-P(y=1)}} \, e^{-\sum_{i=1}^{d} c_i \log \frac{P(w_i|y=1)}{P(w_i|y=0)}}}$$

$$\Rightarrow P(y=1|x) = \frac{1}{1+e^{-\left(\sum_{i=1}^{d} c_i \log\frac{P(w_i|y=1)}{P(w_i|y=0)} + \log\frac{P(y=1)}{1-P(y=1)}\right)}}$$

$$\Rightarrow P(y=1|x) = \frac{1}{1+e^{-\left(\sum_{i=1}^{d} x_i \theta_i + \theta_0\right)}} = \frac{1}{1+e^{-(\theta^T x+\theta_0)}}$$

5

$$P(x|\omega_1) = N(0, I)$$

$$P(x|\omega_2) = N([1 \quad 1]^T, I)$$

$$P(x|\omega_3) = \frac{1}{2} N([0.5 \quad 0.5]^T, I) + \frac{1}{2} N([-0.5 \quad 0.5]^T, I)$$

$$P(\omega_1) = P(\omega_2) = P(\omega_3)$$

الف) احتمال posterior برای $x = [0.3 \quad 0.3]^T$

$$P(x|\omega_1) = \frac{1}{2\pi} e^{-\frac{1}{2}[0.3 \quad 0.3]\begin{bmatrix}0.3\\0.3\end{bmatrix}} = 0.1455$$

$$P(x|\omega_2) = \frac{1}{2\pi} e^{-\frac{1}{2}[-0.7 \quad -0.7]\begin{bmatrix}-0.7\\-0.7\end{bmatrix}} = 0.0975$$

$$P(x|\omega_3) = \frac{1}{4\pi} e^{-\frac{1}{2}[-0.2 \quad -0.2]\begin{bmatrix}-0.2\\-0.2\end{bmatrix}} + \frac{1}{4\pi} e^{-\frac{1}{2}[0.8 \quad -0.2]\begin{bmatrix}0.8\\-0.2\end{bmatrix}} = 0.1331$$

$$P(\omega_i|x) \propto P(x|\omega_i)\, P(\omega_i)$$

ازآنجایی که $P(\omega_1) = P(\omega_2) = P(\omega_3)$ است پس:

$$P(\omega_i|x) \propto P(x|\omega_i)$$

پس چون احتمال $P(x|\omega_1)$ بیشتر از همه است، پس x متعلق به کلاس 1 می‌شود.

ب) $x = [* \quad 0.3]^T$ ($x_1 \quad x_2$)

$$P(\omega_i \mid x_2) = \frac{\int p(x_1, x_2 \mid \omega_i)\, p(\omega_i)\, dx_1}{P(x_2)}$$

$$P(\omega_i \mid x_2) \propto \int p(x_1, x_2 \mid \omega_i) \overset{\text{یکسان}}{p(\omega_i)}\, dx_1$$

چون $p(\omega_i)$ ها برابر است :

$$P(\omega_i \mid x_2) \propto \int p(x_1, x_2 \mid \omega_i)\, dx_1$$

$$\int p(x_1, x_2 \mid \omega_1)\, dx_1 = \frac{1}{2\pi} \int e^{-\frac{1}{2}(x_1^2 + 0.09)}\, dx_1$$

$$= \frac{1}{2\pi} e^{\left(\frac{-0.09}{2}\right)} \int e^{-\frac{1}{2} x_1^2}\, dx_1 =$$

recall that: (integral of gaussian func.)

$$\int_{-\infty}^{+\infty} e^{-a(x+b)^2}\, dx = \sqrt{\frac{\pi}{a}}$$

$$\int p(x_1, x_2 \mid \omega_1)\, dx_1 = \frac{1}{2\pi} e^{\left(\frac{-0.09}{2}\right)} \sqrt{2\pi} = 0.3813$$

$$\int p(x_1, x_2 \mid \omega_2)\, dx_1 = \frac{1}{2\pi} \int e^{-\frac{1}{2}\left((x_1-1)^2 + 0.49\right)}\, dx_1$$

$$= \frac{1}{2\pi} e^{\left(\frac{-0.49}{2}\right)} \sqrt{2\pi} = 0.3122$$

$$\int p(x_1, x_2|w_3)\,dx_1 = \frac{1}{4\pi}\int e^{-\frac{1}{2}\left((x_1-0.5)^2+0.04\right)}\,dx_1$$

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

$$= \frac{1}{4\pi} e^{\left(-\frac{0.04}{2}\right)} \cdot \sqrt{2\pi} + \frac{1}{4\pi} e^{\left(-\frac{0.04}{2}\right)} \sqrt{2\pi}$$

$$= 0.391$$

درنتیجه احتمال کلاس $\underline{3}$ بیشتر از بقیه شد پس $x = [*\ \ 0.3]^T$ مربوط به کلاس $\underline{3}$ است.

---

⑥ آ)
$$J(\theta) = (X\theta - y)^T W (X\theta - y)$$

اگر ماتریس W را شکلی فرض کنیم که عناصر روی قطر اصلی آن $W_{ii} = \frac{1}{2} w^{(i)}$ باشند و بقیه عناصر نیز صفر باشند. همچنین $Z = X\theta - y$ باشد:

$$Z_i = \theta^T x^{(i)} - y^{(i)} \quad , \quad W = \begin{bmatrix} \frac{1}{2}w^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{2}w^{(2)} & & & \\ 0 & & & & \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & \frac{1}{2}w^{(m)} \end{bmatrix} \text{ :آنگاه}$$

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

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

6 ب)

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

$$= 2 X^T W X \theta - 2 X^T W y$$

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

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

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

ج)

$$ML(\theta) = \arg\max_\theta \prod_{i=1}^{m} p\left(y^{(i)} \mid x^{(i)}; \theta\right)$$

$$= \arg\max_\theta \sum_{i=1}^{m} \log p\left(y^{(i)} \mid x^{(i)}; \theta\right)$$

مستقل از θ

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

$$= \arg\max_\theta \sum_{i=1}^{m} - \frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2 \sigma^{(i)2}} \leqslant 0$$ negative or zero

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

فرم linear regression

$$= \arg\min_\theta \frac{1}{2} \sum_{i=1}^{m} \underbrace{\frac{1}{\sigma^{(i)2}}}_{w^{(i)}} \left(y^{(i)} - \theta^T x^{(i)}\right)^2 \Rightarrow \boxed{w^{(i)} = \frac{1}{\sigma^{(i)2}}}$$

Naive Bayes دسته‌بندی 7

$X = [X_1, X_2]^T$ دو ویژگی و $\left.\begin{matrix} c=0 \\ c=1 \\ c=2 \end{matrix}\right\}$ سه کلاس

$$P(x_1 \mid y=0) = Ber(x_1, 0.5) = 0.5^{x_1} \, 0.5^{1-x_1} = 0.5$$

$$P(x_1 \mid y=1) = 0.5 \quad , \quad P(x_1 \mid y=2) = 0.5$$

$$P(x_2 \mid y=0) = Normal(x_2, -1, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}}$$

$$P(x_2 \mid y=1) = Normal(x_2, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

$$P(x_2 \mid y=2) = Normal(x_2, 1, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}}$$

الف) $P(y \mid x_1 = 0, x_2 = 0)$

فرض: $X = [x_1 \;\; x_2]^T$

bayes theorem: $$P(y=c \mid X) = \frac{P(X \mid y=c)\, P(y=c)}{P(X)}$$

ind. Var

$$\Rightarrow P(y=c \mid X) = \frac{\prod_{i=1}^{n} P(x_i \mid y=c) \cdot P(y=c)}{P(X)}$$

Total probability:

$$P(X) = \sum_{c \in C} P(X \mid y=c)\, P(y=c)$$

$$P(X) = P(X|y=0)\,P(y=0) + P(X|y=1)\,P(y=1) + P(X|y=2)\,P(y=2)$$

$$P(X|y=0) = P(x_1|y=0)\,P(x_2|y=0) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2+1)^2}{2}}$$

⊛ $X = [x_1=0 \quad x_2=0]^T$

$$P(X|y=0) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}}$$

$$P(X|y=1) = P(x_1|y=1)\,P(x_2|y=1) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}}$$

$$P(X|y=1) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{0} = \frac{1}{2\sqrt{2\pi}}$$

$$P(X|y=2) = P(x_1|y=2)\,P(x_2|y=2) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2-1)^2}{2}}$$

$$P(X|y=2) = 0.5 \times \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}}$$

$$\Rightarrow P(X) = \frac{1}{4\sqrt{2\pi e}} + \frac{1}{8\sqrt{2\pi}} + \frac{1}{8\sqrt{2\pi e}}$$

$$P(X) = \frac{3}{8\sqrt{2\pi e}} + \frac{1}{8\sqrt{2\pi}} = \frac{3e^{\frac{-1}{2}}+1}{8\sqrt{2\pi}}$$

الف)

$$P(y=0|x) = \frac{P(x|y=0)\cdot P(y=0)}{P(x)} =$$

$$\Rightarrow P(y=0|x) = \frac{\frac{e^{-\frac{1}{2}}}{4\sqrt{2\pi}}}{\frac{3e^{-\frac{1}{2}}+1}{8\sqrt{2\pi}}} = \frac{2e^{-\frac{1}{2}}}{3e^{-\frac{1}{2}}+1}$$

$$P(y=1|x) = \frac{\frac{1}{8\sqrt{2\pi}}}{\frac{3e^{-\frac{1}{2}}+1}{8\sqrt{2\pi}}} = \frac{1}{3e^{-\frac{1}{2}}+1}$$

$$P(y=2|x) = \frac{\frac{e^{-\frac{1}{2}}}{8\sqrt{2\pi}}}{\frac{3e^{-\frac{1}{2}}+1}{8\sqrt{2\pi}}} = \frac{e^{-\frac{1}{2}}}{3e^{-\frac{1}{2}}+1}$$

$$P_{Y|X_1,X_2} = \begin{bmatrix} \frac{2e^{-1/2}}{3e^{-1/2}+1} \\ \frac{1}{3e^{-1/2}+1} \\ \frac{e^{-1/2}}{3e^{-1/2}+1} \end{bmatrix}$$

ب) $P(y | x_1=0)$

$$P(y=0 | x_1=0) = \frac{P(x_1=0 | y=0)\ P(y=0)}{P(x_1=0)}$$

$$P(y=1 | x_1=0) = \frac{P(x_1=0 | y=1)\ P(y=1)}{P(x_1=0)}$$

$$P(y=2 | x_1=0) = \frac{P(x_1=0 | y=2)\ P(y=2)}{P(x_1=0)}$$

$$\begin{cases} P(x_1=0) = \sum\limits_{c \in C} P(x_1=0 | y=c)\ P(y=c) \\ \quad = 0.5 \times 0.5 + 0.5 \times 0.25 + 0.5 \times 0.25 \\ P(x_1=0) = 0.5 \end{cases}$$

$$P(y=0 | x_1=0) = \frac{0.5 \times 0.5}{0.5} = 0.5$$

$$P(y=1 | x_1=0) = \frac{0.5 \times 0.25}{0.5} = 0.25$$

$$P(y=2 | x_1=0) = \frac{0.5 \times 0.25}{0.5} = 0.25$$

$$P_{Y|X_1} = \begin{bmatrix} 0.5 \\ 0.25 \\ 0.25 \end{bmatrix}$$

ii) $P(y \mid x_2=0)$

$$P(x_2=0 \mid y=0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$$

$$P(x_2=0 \mid y=1) = \frac{1}{\sqrt{2\pi}}$$

$$P(x_2=0 \mid y=2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}}$$

$$P(x_2=0) = \sum_{c \in C} P(x_2=0 \mid y=c)\, P(y=c)$$

$$= \frac{e^{-1/2}}{\sqrt{2\pi}} \times 0.5 + \frac{1}{\sqrt{2\pi}} 0.25 + \frac{e^{-1/2}}{\sqrt{2\pi}} 0.25$$

$$P(x_2=0) = \frac{0.75\, e^{-1/2} + 0.25}{\sqrt{2\pi}}$$

$$P(y=0 \mid x_2=0) = \frac{\frac{e^{-1/2}}{\sqrt{2\pi}} \times 0.5}{\frac{0.75\, e^{-1/2} + 0.25}{\sqrt{2\pi}}} = \frac{0.5\, e^{-1/2}}{0.75\, e^{-1/2} + 0.25}$$

$$P(y=1 \mid x_2=0) = \frac{\frac{0.25}{\sqrt{2\pi}}}{\frac{0.75\, e^{-1/2} + 0.25}{\sqrt{2\pi}}} = \frac{0.25}{0.75\, e^{-1/2} + 0.25}$$

$$P(y=2 \mid x_2=0) = \frac{\frac{e^{-1/2}}{\sqrt{2\pi}} \times 0.25}{\frac{0.75\, e^{-1/2} + 0.25}{\sqrt{2\pi}}} = \frac{0.25\, e^{-1/2}}{0.75\, e^{-1/2} + 0.25}$$

پ)

$$P_{Y|X_2} = \begin{bmatrix} \dfrac{0.5\, e^{-1/2}}{0.75\, e^{-1/2} + 0.25} \\[2ex] \dfrac{0.25}{0.75\, e^{-1/2} + 0.25} \\[2ex] \dfrac{0.25\, e^{-1/2}}{0.75\, e^{-1/2} + 0.25} \end{bmatrix}$$

ت) متوجه می‌شویم دو احتمال بخش 'الف' و 'پ' یکسان است:

$$P(y \mid x_1 = 0,\ x_2 = 0) = P(y \mid x_2 = 0)$$

این موضوع بدین دلیل است که $P(x_1=0 \mid y)$ به ازای هر $y=c$ یکسان و برابر با 0.5 است و در بخش الف صرفا احتمال $P(x_1=0 \mid y)$ در صورت و مخرج کسر ضرب شده است که با یک فاکتور گرفتن و حذف کردن آن از صورت و مخرج، به احتمال بخش 'پ' می‌رسیم.

$$P(y \mid x_1 = 0,\ x_2 = 0) = P(y \mid x_2 = 0)$$

$$\frac{P(x_1=0 \mid y=c)\, P(x_2=0 \mid y=c)\, P(y=c)}{P(x_1=0,\ x_2=0)} = \frac{P(x_2=0 \mid y=c)\, P(y=c)}{P(x_2=0)}$$

$$\frac{\overbrace{\cancel{P(x_1=0 \mid y)}}^{\text{Constant مستقل از c}} \times P(x_2=0 \mid y=c)\, P(y=c)}{\underbrace{\cancel{P(x_1=0 \mid y)}}_{\text{Constant}}\ P(x_2=0)} = \frac{P(x_2=0 \mid y=c)\, P(y=c)}{P(x_2=0)}$$