تمرین ۱ ۱ = علیرضا شجاعی ۹۸۱۰۱۸۳۳

۱. Naive Bayes $P(Y=1)=P(Y=2)=P(Y=3)=\frac{1}{3}$

همه احتمال‌ها باهم برابرند پس از مقایسه‌شان در رده‌بندی صرف‌نظر می‌کنیم $P(Y_i|X) = P(X|Y_i)P(Y_i)$

باید ماکسیمم $P(X|Y_i)$ را به دست آورد $i = \arg\max P(X|Y_i) = \arg\max N(\mu_i, \Sigma_i)$

$\mu_1 = [0, 0]^T \quad \mu_2 = [1, 1]^T \quad \mu_3 = [1, 1]^T$

$$\Sigma_1 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 0.2 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

$$\frac{A}{\sqrt{\det \Sigma_i}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \sim N(\mu_i, \Sigma_i)$$

از تابع بالا $\ln$ می‌گیریم که برابر با $\ln A - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - \frac{1}{2}\ln\det\Sigma_i$ می‌شود که A ثابت است. بنابراین برای $\arg\max P(X|Y_i)$ باید کلاسی را پیدا کنیم که $-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) - \frac{1}{2}\ln(\det(\Sigma_i))$ به ازای داده‌ی ورودی داده شده در آن بیشینه شود.

یا معادل آن $(x-\mu)^T \Sigma^{-1}(x-\mu) + 2\ln\det(\Sigma_i)$ کمینه شود.

$$\Sigma_1^{-1} = \begin{bmatrix} 10/7 & 0 \\ 0 & 10/7 \end{bmatrix} \quad \Sigma_2^{-1} = \frac{1}{0.16 - 0.09} \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.8 \end{bmatrix} = \begin{bmatrix} 20/7 & -30/7 \\ -30/7 & 80/7 \end{bmatrix}$$

$$\Sigma_3^{-1} = \frac{1}{0.56 - 0.04} \begin{bmatrix} 0.8 & -0.2 \\ -0.2 & 0.7 \end{bmatrix} = \begin{bmatrix} 80/52 & -20/52 \\ -20/52 & 70/52 \end{bmatrix}$$

$2\ln\det\Sigma_1 = -1.4267 \quad 2\ln\det\Sigma_2 = -1.3078$

$2\ln\det\Sigma_3 = -5.3985$

الف) $x = [50, 0.15]^T$

① $[50 \quad 0.15] \begin{bmatrix} 10/7 & 0 \\ 0 & 10/7 \end{bmatrix} \begin{bmatrix} 50 \\ 0.15 \end{bmatrix} + 2\ln\det\Sigma_1 = [500/7 \quad 1.5/7] \begin{bmatrix} 50 \\ 0.15 \end{bmatrix} + 2\ln\det\Sigma_1 = 3575.59$

② $[49 \quad -0.15] \begin{bmatrix} 20/7 & -30/7 \\ -30/7 & 80/7 \end{bmatrix} \begin{bmatrix} 49 \\ -0.15 \end{bmatrix} + 2\ln\det\Sigma_2 = \left[\frac{990}{7} \quad \frac{-1510}{7}\right] \begin{bmatrix} 49 \\ -0.15 \end{bmatrix} + 2\ln\det\Sigma_2 = 7547.842$

③ $[49 \quad -0.15] \begin{bmatrix} 80/52 & -20/52 \\ -20/52 & 70/52 \end{bmatrix} \begin{bmatrix} 49 \\ -0.15 \end{bmatrix} = \left[\frac{2945}{52} \quad -\frac{1165}{52}\right] \begin{bmatrix} 49 \\ -0.15 \end{bmatrix} + 2\ln\det\Sigma_3 = 3711.741$

پس به کلاس اول تعلق دارد. زیرا کمترین مقدار به ازای آن است. $x \in i=1$

① $[0.5 \quad 0.5] \begin{bmatrix} 2/7 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + 2\ln\det\Sigma_1 = [5/7 \quad 5/2] \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + 2\ln\det\Sigma_1 = -0.113$

② $[-0.5 \quad -0.5] \begin{bmatrix} 20/7 & -30/7 \\ -30/7 & 80/7 \end{bmatrix} \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + 2\ln\det\Sigma_2 = [55/7 \quad -25/7] \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + 2\ln\det\Sigma_2 = -2.4925$

③ $[-0.5 \quad -0.5] \begin{bmatrix} 80/52 & -20/52 \\ -20/52 & 70/52 \end{bmatrix} \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + 2\ln\det\Sigma_3 = [-30/52 \quad -25/52] \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + 2\ln\det\Sigma_3 = -0.7802$

$x \in i=2$

بنابراین مقدار عبارت ~~دوم~~ کمتر است بنابراین x به کلاس ~~دوم~~ برچسب می‌خورد

$$y(x_n,w) = w_0 + \sum_{i=1}^{D} w_i x_i$$ ۲.

$$E_D(w) = \frac{1}{2}\sum_{n=1}^{N}[y(x_n,w)-y_n]^2 \qquad \epsilon_i \sim N(0,\sigma^2 I)$$

$$\Rightarrow y(x_n+\epsilon, w) = w_0 + \sum_{i=1}^{D} w_i(x_i+\epsilon_i) = w_0 + \sum_{i=1}^{D} w_i x_i + \sum_{i=1}^{D} w_i\epsilon_i$$

$$= w_0 + \sum_{i=1}^{D} w_i x_i + \sum_{i=1}^{D} w_i\epsilon_i = y(x_n,w) + \sum_{i=1}^{D} w_i\epsilon_i$$

$$\tilde{E}_D(w) = \frac{1}{2}\sum_{n=1}^{N}\Big[y(x_n,w)-y_n+\sum_{i=1}^{D} w_i\epsilon_i\Big]^2 = \frac{1}{2}\sum_{n=1}^{N}[y(x_n,w)-y_n]^2$$

$$+ \frac{1}{2}\sum_{n=1}^{N}2[y(x_n,w)-y_n]\Big[\sum_{i=1}^{D} w_i\epsilon_i\Big] + \cancel{\frac{1}{2}\sum_{n=1}^{N}\Big(\sum_{i=1}^{D} w_i\epsilon_i\Big)^2} \quad \frac{N}{2}\Big(\sum_{i=1}^{D} w_i\epsilon_i\Big)^2$$

$$\Rightarrow E[\tilde{E}_D(w)] = \frac{1}{2}\sum_{n=1}^{N}[y(x_n,w)-y_n]^2 + \overset{0}{\cancel{E\Big[\sum_{i=1}^{N} w_i\epsilon_i \sum (y(x_n,w)-y_n)\Big]}}$$

$$+ \frac{N}{2}E\Big[\Big(\sum_{i=1}^{D} w_i\epsilon_i\Big)^2\Big] = \frac{N}{2}E\Big[w_1^2\epsilon_1^2+\dots+w_D^2\epsilon_D^2+\cancel{2w_1w_2\epsilon_1\epsilon_2+\dots+2w_{D-1}w_D\epsilon_{D-1}\epsilon_D}\Big]$$

$$\Rightarrow \frac{N}{2}E\Big[\sum w_i^2\epsilon_i^2\Big], \quad E[\epsilon_i^2] = \sigma^2 \Rightarrow \frac{N}{2}E\Big[\sum_{i=1}^{D} w_i^2\epsilon_i^2\Big] = \frac{N}{2}E\Big[\sum_{i=1}^{D} w_i^2\sigma^2\Big]$$

$$= \frac{N}{2}\sigma^2 E\Big[\sum_{i=1}^{D} w_i^2\Big] = \frac{N}{2}\sigma^2 E[w^T w] = \frac{N}{2}\sigma^2 w^T w$$

$$\Rightarrow \tilde{E}_D(w) = E_D(w) + \frac{N\sigma^2}{2}w^T w$$

۴_ الف) اگر یکی از ویژگی های $X_{n\times m}$ که یکی از ستون های ماتریس است را x_j در نظر بگیریم و بخواهیم $\|x_j w_j - y\|^2$ را کمینه کنیم که مقدار بهینه برای پارامتر آن را به دست آوریم، آنگاه داریم:

$$\frac{\partial}{\partial w_j}\|x_j w_j - y\|^2 = 0 \Rightarrow \frac{\partial}{\partial w_j}\sum (x_{ij} w_j - y)^2 = \frac{\partial}{\partial w_j}\sum (x_{ij}^2 w_j^2 + y^2 - 2x_{ij} w_j y) = 0$$

$$\sum 2x_{ij}^2 w_j - \sum 2x_{ij} y = 0 \Rightarrow w_j \sum x_{ij}^2 = \sum x_{ij} y \Rightarrow w_j = \frac{\sum x_{ij} y}{\sum x_{ij}^2} = \frac{x_j^T y}{x_j^T x_j}$$

ب) حال برای حالتی که ماتریس X اگر بخواهیم جواب بهینه برای $XW=y$ به دست آوریم که جواب ما برابر $\underline{W}$ باشد
اگر روش پسین را راستی آن اجرا کنیم که بهینه ترین جواب W که $\|XW-y\|$ را کمینه کند آن جواب $\underline{W}$
به دست زیر خواهد بود.

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

حال $X^T X$ را به دست می‌آوریم برای x_i در x_j به دلیل متعامد بودن ستون‌های این ماتریس برابر صفر خواهد بود.
نقطه به ازای اعضای برابر این مقدار ناصفر است بنابراین نقطه‌های قطر اصلی مقدار ناصفر دارند.

$$\Rightarrow X^T X = \begin{bmatrix} x_1^T x_1 & & & 0 \\ & \ddots & & \\ & & x_j^T x_j & \\ 0 & & & \ddots \; x_m^T x_m \end{bmatrix} \Rightarrow (X^T X)^{-1} = \begin{bmatrix} (x_1^T x_1)^{-1} & & & 0 \\ & \ddots & & \\ & & (x_j^T x_j)^{-1} & \\ 0 & & & (x_m^T x_m)^{-1} \end{bmatrix}$$

$$X^T y = \begin{bmatrix} x_1^T \\ \vdots \\ x_j^T \\ \vdots \\ x_m^T \end{bmatrix} y = \begin{bmatrix} x_1^T y \\ \vdots \\ x_j^T y \\ \vdots \\ x_m^T y \end{bmatrix}$$

$$\Rightarrow \underline{W} = (X^T X)^{-1} X^T y = \begin{bmatrix} (x_1^T x_1)^{-1} & & & 0 \\ & \ddots & & \\ & & (x_j^T x_j)^{-1} & \\ 0 & & & (x_m^T x_m)^{-1} \end{bmatrix} \begin{bmatrix} x_1^T y \\ \vdots \\ x_j^T y \\ \vdots \\ x_m^T y \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_m \end{bmatrix}$$

حال اگر مولفه w_j را در نظر بگیریم:

$$w_j = (x_j^T x_j)^{-1} x_j^T y = \frac{x_j^T y}{x_j^T x_j}$$

همچنین برای هر کدام از مولفه‌ها این عبارت را داریم بنابراین در حالت متعامد تک تک مولفه‌ها با بررسی جداگانه مولفه‌ها در ماتریس کلی یکسان است.

۴-ب) برای آنکه $w = [w_j, w_0]$ را به دست آوریم، ماتریس X را دو ستونی در نظر می‌گیریم
ماتریس X به صورت $X = [x_j \ \ 1]$ به دست می‌آید که ستون 1 مربوط به w_0 و ستون x_j مربوط به w_j می‌باشد.

$$\Rightarrow X^T X = \begin{bmatrix} x_j^T x_j & \sum_{i=1}^{n} x_{ji} \\ \sum_{i=1}^{n} x_{ji} & n \end{bmatrix} \Rightarrow (X^T X)^{-1} = \frac{1}{n \sum_{i=1}^{n} x_{ji}^2 - (\sum_{i=1}^{n} x_{ji})^2} \begin{bmatrix} n & -\sum_i x_{ji} \\ -\sum_i x_{ji} & x_j^T x_j \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} w_j \\ w_0 \end{bmatrix} = (X^T X)^{-1} X^T y, \quad X^T y = \begin{bmatrix} x_j^T y \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} w_j \\ w_0 \end{bmatrix} = (X^T X)^{-1} X^T y = \frac{1}{n \sum_{i=1}^{n} x_{ji}^2 - (\sum_{i=1}^{n} x_{ji})^2} \begin{bmatrix} n & -\sum_i x_{ji} \\ -\sum_i x_{ji} & x_j^T x_j \end{bmatrix} \begin{bmatrix} x_j^T y \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

$$\Rightarrow w_j = \frac{n x_j^T y - \sum_i x_{ji} \sum_i y_i}{n \sum_i x_{ji}^2 - (\sum_i x_{ji})^2} = \frac{\frac{\sum_i x_{ji} y_i}{n} - \frac{\sum x_{ji}}{n}\frac{\sum y_i}{n}}{\frac{\sum x_{ji}^2}{n} - (\frac{\sum x_{ji}}{n})^2} = \frac{E[x_j y] - E[x_j]E[y]}{E[x_j^2] - E[x_j]^2}$$

$$= \frac{\mathrm{cov}(x_j, y)}{\mathrm{var}(x_j)}$$

$$w_0 = \frac{-\sum_i x_{ji}\, x_j^T y + x_j^T x_j \sum y_i}{n \sum_i x_{ji}^2 - (\sum_i x_{ji})^2} = \frac{\sum_i x_{ji}^2 \sum_i y_i - \sum_i x_{ji} \sum_i x_{ji} y_i}{n \sum_i x_{ji}^2 - (\sum_i x_{ji})^2} = \frac{E[x_j^2]E[y_i] - E[x_j]E[x_j y]}{E[x_j^2] - E[x_j]^2}$$

$$= E[y] + \frac{E[x_j]^2 E[y_i] - E[x_j]E[x_j y]}{\mathrm{var}(x_j)} = E[y] - E[x_j]\frac{\mathrm{cov}(x_j, y)}{\mathrm{var}(x_j)} =$$

$$= E[y] - w_j E[x_j]$$

حل آ) $P(X \geq \alpha) \leq \frac{E(X)}{\alpha}$ نامساوی مارکوف

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{0}^{\infty} x f_X(x)\,dx \overset{a>0}{\geq} \int_{a}^{\infty} x f_X(x)\,dx \overset{x \geq a}{\geq} \int_{a}^{\infty} a f_X(x)\,dx$$

$$= a\int_{a}^{\infty} f_X(x)\,dx = a P(X \geq a) \Rightarrow P(X \geq a) \leq \frac{E(X)}{a}, \quad a>0$$

ب) $P(|Z-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$ نامساوی چبیشف $\quad E[Z]=\mu, \ var(Z)=\sigma^2$

$$\xRightarrow{\text{مارکوف}} P(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2}, \quad Y = (Z-E[Z])^2 = (Z-\mu)^2$$

$$E[Y] = E[(Z-\mu)^2] = Var(Z) = \sigma^2$$

$$P(Y \geq \varepsilon^2) = P((Z-\mu)^2 \geq \varepsilon^2) = P(|Z-\mu| \geq \varepsilon) \leq \frac{E(Y)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

ج) تخمین Ω به این صورت خواهد بود تعداد زیادی نقطه را به طور تصادفی داخل مربع پرتاب می‌کنند و تخمین Ω برابر مجموع تعداد متغیرهای تصادفی داخل این دایره به تعداد کل نقاط خواهد بود.

$$\hat{\Omega}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

حال داریم:

$$\boxed{E[X_i] = \Omega \times 1 + (4-\Omega) \times 0 = \Omega}$$

$$var(X_i) = \Omega(4-\Omega)$$

$$\boxed{E[\hat{\Omega}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{n\Omega}{n} = \Omega}$$

$$var(\hat{\Omega}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{(4-\Omega)\Omega}{n^2}$$

اگه با احتمال ۹۵ درصد از ۱ درصد کمتر است به این معنی است که با قطعیت ۵ درصد از ۱ درصد بیشتر باشد.

$$P(|\hat{\Omega}(n) - \Omega| \geq 0.001) \leq 1 - 0.95 = 0.05 \Rightarrow \frac{(4-\Omega)\Omega}{n^2(0.001)^2} \leq 0.05$$

با فرض عدم اطلاع از Ω، آن را طوری تعیین می‌کنیم که به ازای آن n ماکسیمم شود

$$\Rightarrow n \geq \frac{(4-\Omega)\Omega}{0.05\,(0.001)^2} = A$$

$$\Rightarrow \frac{dA}{d\Omega} = 0 \Rightarrow \Omega = 2 \Rightarrow \boxed{n \geq \frac{(4-2)2}{0.05\,(0.001)^2} = 80{,}000{,}000}$$

NAHAL

Scanned with CamScanner

۶ـ ۰

$\sigma_{max}(A) = \|A\|_2 =$ بزرگترین مولفه Σ

$\sigma_{max}(A^{-1}) = \|A^{-1}\|_2 =$ بزرگترین مولفه Σ^{-1}

از آن جایی که وقتی ماتریس A معکوس می‌شود، مقادیر تکین آن هم معکوس می‌شود بنابراین $\sigma_{max}(A)$ تبدیل به $1/\sigma_{max}(A)$

در معکوس A می‌شود که برابر کوچکترین مقدار تکین A^{-1} خواهد بود بنابراین بزرگترین مقدار تکین A^{-1} نیز برابر $1/\sigma_{min}(A)$

می‌باشد حتماً $1/\sigma_{min}(A)$ بزرگتر خواهد بود از $1/\sigma_{max}(A)$ $\Rightarrow 1 \leq \sigma_{max}(A)\sigma_{max}(A^{-1})$ $\quad \frac{1}{\sigma_{max}(A)} \leq \sigma_{max}(A^{-1})$

۰

$$\left.\begin{aligned} \|A\|_2^2 &= \sigma_1^2 \\ \|A\|_F^2 &= \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2 \\ \operatorname{rank}(A) &= r \Rightarrow \operatorname{rank}(A)\|A\|_2^2 = r\sigma_r^2 \end{aligned}\right\} \qquad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$$

$$\Rightarrow \|A\|_2^2 \leq \|A\|_F^2 \leq \operatorname{rank}(A)\|A\|_2^2$$

$$\Rightarrow \|A\|_2 \leq \|A\|_F \leq \sqrt{\operatorname{rank}(A)}\,\|A\|_2$$

٧ـ

$$y(x,w) = w_0 + \sum_{j=1}^n \left(w_j\, \sigma\left(2\frac{x-\mu_j}{s}\right)\right) \qquad \sigma(x) = \frac{1}{1+e^{-x}}$$

$$y(x,u) = u_0 + \sum_{j=1}^n \left[u_j \tanh\left(\frac{x-\mu_j}{s}\right)\right] \qquad \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$y(x,w) = w_0 + \sum_{j=1}^n w_j \frac{1}{1+e^{2\frac{\mu_j - x}{s}}}$$

$$y(x,u) = u_0 + \sum_{j=1}^n \left[u_j \tanh\left(\left(\frac{x-\mu_j}{s}\right)\right)\right] = u_0 + \sum_{j=1}^n u_j \frac{e^{\frac{x-\mu_j}{s}} - e^{-\frac{x-\mu_j}{s}}}{e^{\frac{x-\mu_j}{s}} + e^{-\frac{x-\mu_j}{s}}}$$

$$\Rightarrow y(x,u) = u_0 + \sum u_j \frac{1 - e^{-2\frac{x-\mu_j}{s}}}{1 + e^{-2\frac{x-\mu_j}{s}}} = u_0 + \sum u_j \frac{2 - \left(1 + e^{-2\frac{x-\mu_j}{s}}\right)}{1 + e^{-2\frac{x-\mu_j}{s}}}$$

$$= u_0 + \sum u_j \left(\frac{2}{1+e^{-2\frac{x-\mu_j}{s}}} - 1\right) = u_0 + \sum_{j=1}^n \left[2u_j \frac{1}{1+e^{-2\frac{x-\mu_j}{s}}} - u_j\right]$$

$$= \left(u_0 - \sum_{j=1}^n u_j\right) + \sum_{j=1}^n 2u_j\, \sigma\left(2\frac{x-\mu_j}{s}\right) \Rightarrow \begin{cases} w_0 = u_0 - \sum_{j=1}^n u_j \\ w_j = 2u_j \end{cases}$$

Scanned with CamScanner