Теоретическое д/з. (18.11.22)

№1.

Д-ть, что:

1) если $a \in R^n$, $x \in R^n \Rightarrow \frac{\partial(a^T x)}{\partial x} = a$

$$a^T x = [a_1\ a_2 \dots a_n] \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (a_1 x_1 + a_2 x_2 + \dots + a_n x_n) = \sum_{i=1}^{n} a_i x_i$$

$$\forall j: \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n} a_i x_i\right) = a_j \Rightarrow \frac{d(\sum a_i x_i)}{dx} \overset{def}{=} \begin{bmatrix} \frac{\partial(\sum a_i x_i)}{\partial x_1} \\ \frac{\partial(\sum a_i x_i)}{\partial x_2} \\ \vdots \\ \frac{\partial(\sum a_i x_i)}{\partial x_n} \end{bmatrix} =$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a \quad \text{(чтд)}$$

2) если $A \in R^{m \times n}$, $x \in R^n \Rightarrow \frac{\partial(Ax)}{\partial x} = A$.

$$Ax = \underset{m \times n}{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}} \times \underset{n \times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} = \underset{m \times 1}{\begin{bmatrix} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n \\ a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \end{bmatrix}} =$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i} x_i \\ \sum a_{2i} x_i \\ \dots \\ \sum a_{mi} x_i \end{bmatrix}$$

$$\forall j: \quad \frac{\partial(Ax)}{\partial x_j} = \begin{bmatrix} \frac{\partial}{\partial x_j}(\sum a_{1i} x_i) \\ \dots \\ \frac{\partial}{\partial x_j}(\sum a_{mi} x_i) \end{bmatrix} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

$$\frac{\partial(Ax)}{\partial x} = \begin{bmatrix} \frac{\partial(\sum a_{1i}x_i)}{\partial x_1} & \frac{\partial(\sum a_{1i}x_i)}{\partial x_2} & \dots & \frac{\partial(\sum a_{1i}x_i)}{\partial x_n} \\ \dots \\ \frac{\partial(\sum^n a_{mi}x_i)}{\partial x_1} & \frac{\partial(\sum^n a_{mi}x_i)}{\partial x_2} & \dots & \frac{\partial(\sum^n a_{mi}x_i)}{\partial x_n} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & \dots & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = A \quad (\text{ч.т.д.})$$

3) если $A \in R^{n \times n}$, $x \in R^n \Rightarrow \frac{\partial(x^T A x)}{\partial x} = (A + A^T)x$

если $A^T = A$, то $\frac{\partial(X^T A x)}{\partial x} = 2Ax$

$$x^T A = [x_1 \; x_2 \dots x_n] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = [x_1 \cdot a_{11} + x_2 \cdot a_{21} + \dots + x_n a_{n1} \; \dots$$

$$\dots \; x_1 \cdot a_{1n} + x_2 \cdot a_{2n} + \dots + x_n a_{nn}] = \left[\sum_{i=1}^n x_i a_{i1} \; \sum_{i=1}^n x_i a_{i2} \dots \sum_{i=1}^n x_i a_{in}\right]$$

$$x^T A x = \left[\sum x_i a_{i1} \; \sum x_i a_{i2} \dots \sum x_i a_{in}\right] \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= \left[x_1 \sum x_i a_{i1} + x_2 \sum x_i a_{i2} + \dots + x_n \sum x_i a_{in}\right]$$

$$\frac{\partial(x^T A x)}{\partial x} = \begin{bmatrix} \frac{\partial(x^T A x)}{\partial x_1} \\ \frac{\partial(X^T A x)}{\partial x_2} \\ \vdots \\ \frac{\partial(X^T A x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \left(\sum x_i a_{i1} + x_1 \cdot a_{11}\right) + x_2 \cdot a_{12} + \dots + x_n a_{1n} \\ x_1 a_{21} + \left(\sum x_i a_{i2} + x_2 \cdot a_{22}\right) + \dots + x_n a_{2n} \\ \\ x_1 a_{n1} + x_2 a_{n2} + \dots + x_n \sum x_i a_{in} + x_n a_{nn} \end{bmatrix}$$

$x_1 a_{11} + x_2 a_{21} + x_3 a_{31} + \dots + x_n a_{n1}$

$$= \begin{bmatrix} \sum x_i a_{i1} + x_1 a_{11} + x_2 \cdot a_{12} + \dots + x_n a_{1n} \\ \sum x_i a_{i2} + x_1 a_{21} + x_2 a_{22} + \dots + x_n a_{2n} \\ \vdots \\ \sum x_i a_{in} + x_1 a_{n1} + x_2 a_{n2} + \dots + x_n a_{nn} \end{bmatrix} =$$

$$= \left(\begin{bmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{n2} \\ a_{1n} & a_{2n} & \dots & & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= (A^T + A)x \quad \text{ч.т.д.}$$

4) если $x \in R^n$, то $\frac{\partial \|x\|^2}{\partial x} = 2x$

$$\|x\|^2 = (x, x) = \sum x_i^2$$

$$\frac{\partial (\sum x_i^2)}{\partial x} = \begin{bmatrix} \frac{\partial \sum x_i^2}{\partial x_1} \\ \frac{\partial \sum x_i^2}{\partial x_2} \\ \vdots \\ \frac{\partial \sum x_i^2}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2x. \quad \text{(чтд)}$$

5) g - скалярная ф-ция; $g(x)$ - применение g к каждой компоненте вектора $x \in R^n$, то $\frac{\partial g(x)}{\partial x} = diag(g'(x))$

$$\forall j: \frac{\partial g(x_j)}{\partial x_j} = g'(x_j)$$

$$\frac{\partial g(x)}{\partial x} \overset{def}{=} \begin{bmatrix} \frac{\partial g(x_1)}{\partial x_1} & \frac{\partial g(x_2)}{\partial x_1} & \dots & \frac{\partial g(x_n)}{\partial x_1} \\ \frac{\partial g(x_2)}{\partial x_2} & \frac{\partial g(x_2)}{\partial x_2} & \dots & \frac{\partial g(x_n)}{\partial x_n} \\ \frac{\partial g(x_n)}{\partial x_n} & \frac{\partial g(x_n)}{\partial x_2} & \dots & \frac{\partial g(x_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} g'(x_1) & 0 & \dots & 0 \\ 0 & g'(x_2) & \dots & 0 \\ 0 & 0 & \dots & g'(x_n) \end{bmatrix} = diag(g'(x))$$

чтд.

6) если $h: R^n \to R^m$, $g: R^m \to R^p$, $x \in R^n$, то

$$\frac{\partial g(h(x))}{\partial x} = \frac{\partial g(h(x))}{\partial h} \frac{\partial h(x)}{\partial x}.$$

$$\frac{\partial g(h(x))}{\partial x} = \begin{bmatrix} \frac{\partial g_1(h(x))}{\partial x_1} & \dots & \frac{\partial g_1(h(x))}{\partial x_n} \\ \frac{\partial g_2(h(x))}{\partial x_1} & \dots & \frac{\partial g_2(h(x))}{\partial x_n} \\ \vdots & & \\ \frac{\partial g_p(h(x))}{\partial x_1} & \dots & \frac{\partial g_p(h(x))}{\partial x_n} \end{bmatrix}$$

$$\forall j: \frac{\partial g_j(h(x))}{\partial x_j} = \frac{\partial g_j(h(x))}{\partial h_1} \cdot \frac{\partial h_1(x)}{\partial x_j} + \frac{\partial g_j(h(x))}{\partial h_2} \cdot \frac{\partial h_2(x)}{\partial x_j} + \dots + \frac{\partial g_j(x)}{\partial h_m} \cdot \frac{\partial h_m(x)}{\partial x_j}$$

$$\Downarrow \quad \underbrace{\begin{bmatrix} \frac{\partial g_1(h(x))}{\partial h_1} & \frac{\partial g_1(h(x))}{\partial h_2} & \dots & \frac{\partial g_1(h(x))}{\partial h_m} \\ \frac{\partial g_p(h(x))}{\partial h_1} & \frac{\partial g_p(h(x))}{\partial h_2} & \dots & \frac{\partial g_p(h(x))}{\partial h_m} \end{bmatrix}}_{= \frac{\partial g(h(x))}{\partial h}}^{p \times m} \cdot \overbrace{\begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \dots & \frac{\partial h_m(x)}{\partial x_n} \\ \frac{\partial h_2(x)}{\partial x_1} & & \vdots \\ \vdots & & \\ \frac{\partial h_m(x)}{\partial x} & & \frac{\partial h_m(x)}{\partial x_n} \end{bmatrix}}^{m \times n} = \frac{\partial h(x)}{\partial x}$$

$$= \frac{\partial g(h)}{\partial h} \cdot \frac{\partial h(x)}{\partial x}$$

№3

$$f(x) = B_0 + B_1 x + B_2 x^2$$

1)

x	1	1	0	0	-1
y	4	4	0	2	6

$$X * B = Y \Rightarrow X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} B_0 \\ B_1 \\ B_2 \end{bmatrix},$$

$$Y = \begin{bmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{bmatrix}$$

Вычислим $X^T X = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix}, \quad X^T Y = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$

Решим систему нормальных ур-ий $X^T X . B = X^T Y$

$$\begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} \times \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix} \Rightarrow \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} \Rightarrow$$

Наша модель $1 - x + 4x^2$

2) Метод ридж-регрессии; $\lambda = 1$ - параметр регуляризации

~~Решим~~ систему $X^T X + \lambda I = X^T X + I = \begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix}$

Наша система $(X^T X + \lambda I) B = X^T Y$

$$\begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix} \Rightarrow \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 3/2 \\ -1/2 \\ 5/2 \end{pmatrix}$$

Получим модель $\frac{3}{2} - \frac{1}{2} x + \frac{5}{2} x^2$

№9

x_1	0	1	0	2	2	2	4	3
x_2	-1	0	0	0	1	0	1	2
y	0	0	0	0	0	1	1	1

1) Оценим вероятности классов:

$$\hat{P}_r\{Y=0\} = \frac{5}{8}, \quad \hat{P}_r\{Y=1\} = \frac{3}{8}$$

Средние для классов: $\hat{\mu}_0 = \begin{pmatrix}1\\0\end{pmatrix}$, $\hat{\mu}_1 = \begin{pmatrix}3\\1\end{pmatrix}$

Объекты распределены по нормальному закону:

$$p(x|y) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma_y}}\, e^{-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1}(x-\mu_y)}$$

мат. ожидание в d мерном пр-ке

$\Sigma_y = \mathrm{cov}\, X$; $\mu_y \in \mathbb{R}^d$

м. ковариации

$\mathrm{cov}(X_i, X_j) = E(X_i - EX_i)(X_j - EX_j)$

центрир.

$\mathrm{cov}(X_i, X_i)$ = дисперсия.

Мат. ковариации для каждого класса:

$$\hat{\Sigma}_0 = \frac{1}{N_0 - 1}\sum_{y^{(i)}=0}(x^{(i)} - \hat{\mu}_0)(x^{(i)} - \hat{\mu}_0)^T = \frac{1}{5-1}\left(\left(\begin{pmatrix}0\\-1\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix}\right)\left(\begin{pmatrix}0\\-1\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix}\right)^T + \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix}\right) \dots\right) =$$

$$= \frac{1}{4}\sum\left(\begin{pmatrix}1&1\\1&1\end{pmatrix} + \begin{pmatrix}0&0\\0&0\end{pmatrix} + \begin{pmatrix}1&0\\0&0\end{pmatrix} + \begin{pmatrix}1&0\\0&0\end{pmatrix} + \begin{pmatrix}1&1\\1&1\end{pmatrix}\right) = \frac{1}{4}\begin{pmatrix}4&2\\2&2\end{pmatrix}$$

$$\hat{\Sigma}_1 = \frac{1}{N_1 - 1}\sum_{y^{(i)}=1}(x^{(i)} - \hat{\mu}_1)(x^{(i)} - \hat{\mu}_1)^T = \frac{1}{2}\left(\begin{pmatrix}1&+1\\+1&1\end{pmatrix} + \begin{pmatrix}1&0\\0&0\end{pmatrix} + \begin{pmatrix}0&0\\0&1\end{pmatrix}\right) = \frac{1}{2}\cancel{\begin{pmatrix}2&2\\2&3\end{pmatrix}} = \frac{1}{2}\begin{pmatrix}2&1\\1&2\end{pmatrix}$$

Оценим матрицу коварияции

$$\hat{\Sigma} = \frac{1}{N-K}\sum_{K}\sum_{y^{(i)}=k}(x^{(i)} - \hat{\mu}_k)(x^{(i)} - \hat{\mu}_k)^T = \frac{1}{6}\left(\begin{pmatrix}4&2\\2&2\end{pmatrix} + \begin{pmatrix}2&2\\2&3\end{pmatrix}\right) =$$

$$= \frac{1}{6}\begin{pmatrix}6&4\\4&5\end{pmatrix} = \begin{pmatrix}1&2/3\\2/3&5/6\end{pmatrix}$$

Найдем обратные:

$$\hat{\Sigma}_0^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}, \quad \hat{\Sigma}_1^{-1} = \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix}, \quad \hat{\Sigma}^{-1} = \begin{pmatrix} 8/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix}$$

Линейные дискриминантные ф-ии

$$\delta_0(x) = x^T \hat{\Sigma}^{-1} \hat{\mu}_0 - \frac{1}{2} \hat{\mu}_0^T \hat{\Sigma}^{-1} \hat{\mu}_0 + \ln \hat{Pr}\{Y=0\} =$$

$$= (x_1\ x_2) \underbrace{\begin{pmatrix} 8/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{= \begin{pmatrix} 8/5 \\ -6/5 \end{pmatrix}} - \frac{1}{2} (1\ 0) \underbrace{\begin{pmatrix} 8/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{= \begin{pmatrix} 8/5 \\ -6/5 \end{pmatrix}} + \ln \frac{5}{8} =$$

$$= \frac{8}{5} x_1 - \frac{6}{5} x_2 - \frac{1}{2} \cdot \frac{8}{5} + \ln \frac{5}{8} = \frac{8}{5} x_1 - \frac{6}{5} x_2 - \frac{4}{5} + \ln \frac{5}{8}$$

$$\delta_1(x) = x^T \hat{\Sigma}^{-1} \hat{\mu}_1 - \frac{1}{2} \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 + \ln \hat{Pr}\{Y=1\} =$$

$$= (x_1\ x_2) \underbrace{\begin{pmatrix} 8/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix}}_{= \begin{pmatrix} 18/5 \\ -6/5 \end{pmatrix}} - \frac{1}{2} (3\ 1) \underbrace{\begin{pmatrix} \ldots \end{pmatrix}}_{= \begin{pmatrix} 18/5 \\ -6/5 \end{pmatrix}} + \ln \frac{3}{8} =$$

$$= \frac{18}{5} x_1 - \frac{6}{5} x_2 - \frac{24}{5} + \ln \frac{3}{8}.$$

$\Downarrow$ Разделяющая поверхность – прямая с ур-ием

$$\delta_0(x) = \delta_1(x) \Rightarrow 2x_1 - 4 + \ln \frac{3}{8} - \ln \frac{5}{8} = 0$$

2) Квадратичные дискриминантные ф-ии:

$$\delta_0(x) = -\frac{1}{2} \ln \det \hat{\Sigma}_0 - \frac{1}{2} (x - \hat{\mu}_0)^T \hat{\Sigma}_0^{-1} (x - \hat{\mu}_0) + \ln \hat{Pr}\{Y=0\} =$$

$$= -\frac{1}{2} \ln 4 - \frac{1}{2} (x_1 - 1\ \ x_2) \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix} + \ln \frac{5}{8} =$$

$$= -\frac{1}{2} \ln 4 + \ln \frac{5}{8} - (x_1 - 1\ \ x_2) \begin{pmatrix} x_1 - 1 - x_2 \\ -x_1 + 1 + 2x_2 \end{pmatrix} = -\frac{1}{2} \ln 4 + \ln \frac{5}{8} - \big((x_1 - 1)(x_1 - x_2 - 1) +$$

$$+ x_2(-x_1 + 2x_2 + 1)\big) = -\frac{1}{2} \ln 4 + \ln \frac{5}{8} - \left(\underline{x_1^2} - 2x_1 x_2 \underline{- 2x_1} + 2x_2 \underline{+ 1} + 2x_2^2\right) =$$

$$= -(x_1 - 1)^2 - 2(x_2^2 - x_1 x_2 + x_2)$$

$$\delta_1(x) = -\frac{1}{6}(4x_1^2 + 4x_2^2 - 4x_1 x_2 - 20x_1 + 4x_2 + 28) - \frac{1}{2} \ln \frac{12}{3} + \ln \frac{3}{8}$$

$\Downarrow$

Разделяющая поверхность:

$$\delta_0(x) = \delta_1(x) = \frac{1}{3} x_1^2 - \frac{4}{3} x_2^2 - \frac{4}{3} x_1 x_2 + \frac{4}{3} x_1 + \frac{4}{3} x_2 - \frac{11}{3} - \frac{1}{2} \ln \frac{12}{3} + \ln \frac{3}{8} +$$

$$+ \frac{1}{2} \ln 4 - \ln \frac{5}{8}.$$

№15

x_1	0	0	1	1	0	0	1	1	1	0
x_2	0	1	0	1	1	1	1	1	1	1
y	0	0	0	0	0	1	1	1	1	1

$Pr(Y=0 \mid X_1=1, X_2=1)$, $Pr(Y=1 \mid X_1=1, X_2=1)$

Оценим априорные вероятности: $\hat{Pr}\{Y=0\} = \frac{1}{2}$, $\hat{Pr}\{Y=1\} = \frac{1}{2}$

Оцениваем условные вероятности:

$\hat{Pr}\{X_1=0 \mid Y=0\} = \frac{3}{5}$, $\hat{Pr}\{X_1=1 \mid Y=0\} = \frac{2}{5}$

$\hat{Pr}\{X_1=0 \mid Y=1\} = \frac{2}{5}$, $\hat{Pr}\{X_1=1 \mid Y=1\} = \frac{3}{5}$

$\hat{Pr}\{X_2=0 \mid Y=0\} = \frac{2}{5}$, $\hat{Pr}\{X_2=1 \mid Y=0\} = \frac{3}{5}$

$\hat{Pr}\{X_2=0 \mid Y=1\} = 0$, $\hat{Pr}\{X_2=1 \mid Y=1\} = 1$.

$$Pr\{Y=0 \mid X_1=1, X_2=1\} = \frac{Pr\{X_1=1 \mid Y=0\}\, Pr\{X_2=1 \mid Y=0\}\, Pr\{Y=0\}}{Pr\{X_1=1, X_2=1\}} =$$

$$= \frac{\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{2}{7}$$

$$= \sum^{k} Pr(Y=k)\, Pr(X=x \mid Y=k)$$

$$Pr\{Y=1 \mid X_1=1, X_2=1\} = \frac{Pr\{X_1=1 \mid Y=1\}\, Pr\{X_2=1 \mid Y=1\}\, Pr\{Y=1\}}{Pr\{X_1=1, X_2=1\}} = \frac{5}{7}$$