сификации, в каждой по $x \in X$ требуется предсказать $y_k \in \{0, 1\}$ $(k = 1, 2, \ldots, K)$. Обучающая выборка составлена из пар $(x^{(i)}, y^{(i)})$, где $y^{(i)} = (y_1^{(i)}, y_2^{(i)}, \ldots, y_K^{(i)}) \in \{0, 1\}^K$ $(i = 1, \ldots, N)$

Для решения исп-ся нейронная сеть, последний слой который вычисляет K функций $g_k(s_k) = \frac{1}{1 + e^{-s_k}}$ $(k = 1, 2, \ldots, K)$

Как штраф используется $R^{(i)} = -\sum_{k=1}^{K} (y_k^{(i)} \ln g_k + (1 - y_k^{(i)}) \ln(1 - g_k))$

Доказать $\frac{\partial R^{(i)}}{\partial s_k} = g_k - y_k^{(i)}$

Д-во:

$$\frac{\partial R^{(i)}}{\partial s_k} = \sum_{k=1}^{K} \frac{\partial R^{(i)}}{\partial g_k} \cdot \frac{\partial g_k}{\partial s_k} \ominus$$

$$\frac{\partial R^{(i)}}{\partial g_k} = -\left(\frac{y_k^{(i)}}{g_k} - \frac{1 - y_k^{(i)}}{1 - g_k}\right)$$

$$\frac{\partial g_k}{\partial s_k} = \frac{e^{-s_k}}{(1 + e^{-s_k})^2} = \frac{e^{-s_k} + 1 - 1}{(1 + e^{-s_k})^2} = \frac{1}{e^{-s_k} + 1} - \frac{1}{(1 + e^{-s_k})^2} =$$

$$= g_k - g_k^2 = g_k(1 - g_k)$$

$$\ominus -\left(\frac{y_k^{(i)}}{g_k} - \frac{1 - y_k^{(i)}}{1 - g_k}\right) \cdot g_k(1 - g_k) = -\frac{y_k^{(i)} g_k (1 - g_k)}{g_k} +$$

$$+ \frac{(1 - y_k^{(i)}) \cdot g_k (1 - g_k)}{1 - g_k} = -y_k^{(i)} \cdot (1 - g_k) + g_k (1 - y_k^{(i)}) =$$

$$= -y_k^{(i)} + y_k^{(i)} \cdot g_k + g_k - g_k \cdot y_k^{(i)} = g_k - y_k^{(i)}$$

1) $I(k=l) = \begin{cases} 1, & k=l \\ 0, & k \neq l \end{cases}$

• $k=l$: $\dfrac{\partial g_k}{\partial s_l} = \left(\dfrac{e^{s_k}}{\sum\limits_{i=1}^{k} e^{s_i}}\right)'_{s_l} = \dfrac{e^{s_k}\left(\sum\limits_{i=1}^{k} e^{s_i}\right) - e^{s_k}\cdot e^{s_l}}{\left(\sum\limits_{i=1}^{k} e^{s_i}\right)^2}$

$= \underbrace{\dfrac{e^{s_k}}{\sum\limits_{i=1}^{k} e^{s_i}}}_{g_k} - \underbrace{\dfrac{e^{s_k}}{\sum\limits_{i=1}^{k} e^{s_i}}}_{g_k} \cdot \underbrace{\dfrac{e^{s_l}}{\sum\limits_{i=1}^{k} e^{s_i}}}_{g_l} = g_k(1-g_l)$

• $k \neq l$: $\dfrac{\partial g_k}{\partial s_l} = \left(\dfrac{e^{s_k}}{\sum\limits_{i=1}^{k} e^{s_i}}\right)'_{s_l} = \dfrac{0 - e^{s_k}\cdot e^{s_l}}{\left(\sum\limits_{i=1}^{k} e^{s_i}\right)^2} =$

$= 0 - \underbrace{\dfrac{e^{s_k}}{\sum\limits_{i=1}^{k} e^{s_i}}}_{g_k} \cdot \underbrace{\dfrac{e^{s_l}}{\sum\limits_{i=1}^{k} e^{s_i}}}_{g_l} = 0 - g_k g_l = -g_k \cdot g_l$

$\Rightarrow \dfrac{\partial g_k}{\partial s_l} = \begin{cases} g_k(1-g_l), & k=l \\ -g_k \cdot g_l, & k \neq l \end{cases} \Rightarrow \dfrac{\partial g_k}{\partial s_l} = g_k(I(k=l)-g_l)$

2) $\dfrac{\partial R^{(i)}}{\partial g_k} = \left(-I(y^{(i)}=k)\ln g_k(s_1,\dots,s_k)\right)'_{g_k} = -I(y^{(i)}=k)\cdot$
$\left(\ln g_k(s_1,\dots,s_k)\right)'_{g_k} = -\dfrac{I(y^{(i)}=k)}{g_k}$

3) $\dfrac{\partial R^{(i)}}{\partial s_l} = \sum\limits_{k=1}^{k} \dfrac{\partial R^{(i)}}{\partial g_k}\cdot\dfrac{\partial g_k}{\partial s_l} = -\sum\limits_{k=1}^{k}\left(\dfrac{I(y^{(i)}=k)}{g_k}\cdot\right.$
$\left. g_k(I(k=l)-g_l)\right) = -\sum\limits_{i=1}^{k}\left(I(y^{(i)}=k)(I(k=l)-g_l)\right)$
$= -(I(y^{(i)}=l) - g_i \cdot I(y^{(i)}=k)) = g_l - I(l=y^{(i)})$

№20

Рассматривается к задач двухклассовой клас-

№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

1) $Pr(Y=0 \mid x_1=1, x_2=1)$
2) $Pr(Y=1 \mid x_1=1, x_2=1)$

$\widehat{Pr}\{Y=0\} = \frac{1}{2}$
$\widehat{Pr}\{Y=1\} = \frac{1}{2}$

$\widehat{Pr}\{x_1=1 \mid Y=0\} = \frac{2}{5}$, $\widehat{Pr}\{x_1=1 \mid Y=1\} = \frac{3}{5}$
$\widehat{Pr}\{x_2=1 \mid Y=0\} = \frac{3}{5}$; $\widehat{Pr}\{x_2=1 \mid Y=1\} = 1$

$$\widehat{Pr}\{Y=0 \mid x_1=1, x_2=1\} = \frac{Pr\{x_1=1 \mid Y=0\} \cdot Pr\{x_2=1 \mid Y=0\} \cdot}{Pr\{x_1=1, x_2=1\}}$$

$$\cdot \underline{Pr\{Y=1\}} = \frac{\frac{3}{5} \cdot 1 \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{\frac{3}{10}}{\frac{21}{50}} = \frac{\cancel{3} \cdot \cancel{50}}{\cancel{10} \cdot \cancel{21}_7} = \frac{5}{7}$$

№19

В задаче классификации на K классов $\{1, 2, \dots, K\}$ последний слой нейронной сети вычисляет softmax-функцию:

$$g_K(s_1, s_2, \dots, s_K) = \frac{e^{s_K}}{\sum_{\ell=1}^{K} e^{s_\ell}}$$

Как штраф использ-ся logloss-функция:

$$R^{(i)} = -\sum_{i=1}^{K} I(y^{(i)} = K) \ln g_K(s_1, s_2, \dots, s_K),$$ где $g_K(s_1, \dots, s_K)$ — softmax-функц.

Д-во:

$$= -\frac{1}{2}\ln\frac{1}{4} - x_1^2 + 2x_1 + 2x_1x_2 - 1 - 2x_2 - 2x_2^2 + \ln\frac{5}{8}$$

$$= -x_1^2 + 2x_1 + 2x_1x_2 - 2x_2 - 2x_2^2 + \ln\frac{5}{4} - 1$$

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

$$= -\frac{1}{2}\cdot\ln\frac{3}{4} - \frac{1}{2}(x_1-3 \quad x_2-1)\begin{pmatrix}\frac{4}{3}x_1 - \frac{2}{3}x_2 - \frac{10}{3} \\ -\frac{2}{3}x_1 + \frac{4}{3}x_2 + \frac{2}{3}\end{pmatrix} + \ln\frac{3}{8}$$

$$= -\frac{1}{2}\cdot\ln\frac{3}{4} - \frac{1}{2}\Big(\frac{4}{3}x_1^2 + \frac{4}{3}x_2^2 - \frac{4}{3}x_1x_2 - \frac{20}{3}x_1 + \frac{4}{3}x_2 + \frac{28}{3}\Big) + \ln\frac{3}{8} =$$

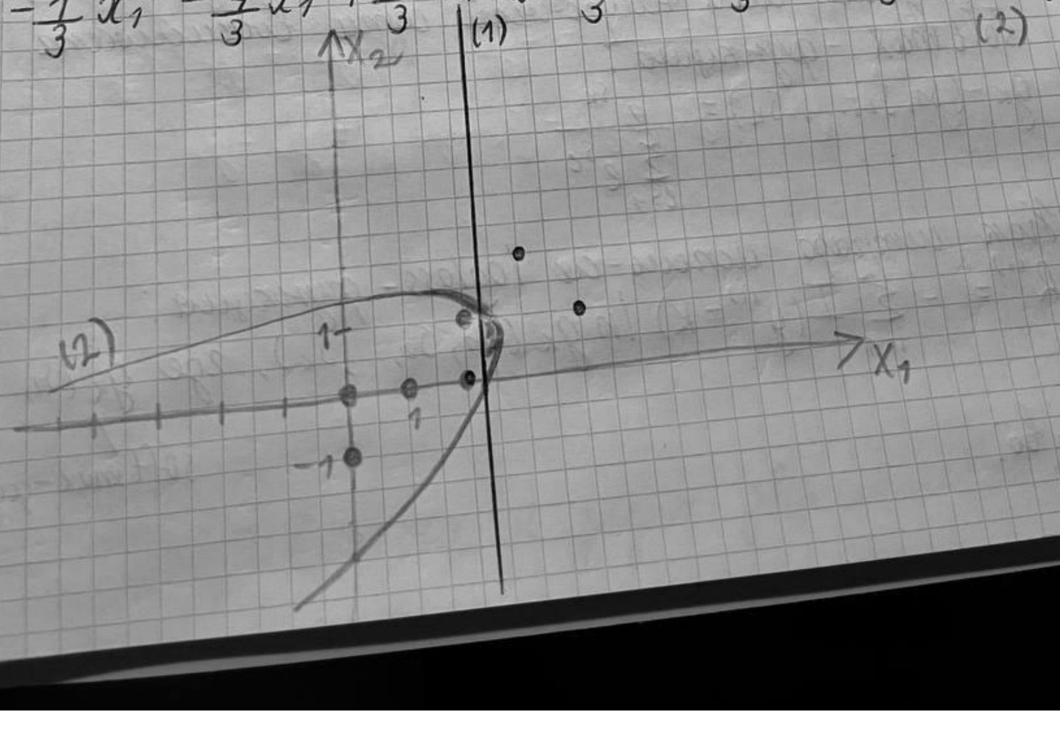
$$= -\frac{1}{2}\ln\frac{3}{4} - \frac{2}{3}x_1^2 - \frac{2}{3}x_2^2 + \frac{2}{3}x_1x_2 + \frac{10}{3}x_1 - \frac{2}{3}x_2 - \frac{14}{3} + \ln\frac{3}{8}$$

Разделяющая поверхность:

$$\delta_0(x) = \delta_1(x)$$

$$-x_1^2 + 2x_1 + 2x_1x_2 - 2x_2 - 2x_2^2 + \ln\frac{5}{4} - 1 = -\frac{1}{2}\ln\frac{3}{4} - \frac{2}{3}x_1^2 - \frac{2}{3}x_2^2 + \frac{2}{3}x_1x_2 + \frac{10}{3}x_1 - \frac{2}{3}x_2 - \frac{14}{3} + \ln\frac{3}{8}$$

$$-\frac{1}{3}x_1^2 - \frac{4}{3}x_1 + \frac{4}{3}x_1x_2 - \frac{4}{3}x_2^2 - \frac{4}{3}x_2 + \frac{11}{3} + \frac{1}{2}\ln\frac{25}{3} = 0$$



Линейные дискриминантные функции:

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

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

$+ \ln\frac{5}{8} =$

$= (x_1 x_2)\cdot\begin{pmatrix} \frac{8}{5} \\ -\frac{6}{5} \end{pmatrix} - \frac{1}{2}(1\ 0)\begin{pmatrix} \frac{8}{5} \\ -\frac{6}{5} \end{pmatrix} + \ln\frac{5}{8} = \frac{8}{5}x_1 - \frac{6}{5}x_2 - \frac{4}{5}$

$+ \ln\frac{5}{8}$

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

$= (x_1 x_2)\begin{pmatrix} \frac{18}{5} \\ -\frac{6}{5} \end{pmatrix} - \frac{1}{2}(3\ \ 1)\begin{pmatrix} \frac{18}{5} \\ -\frac{6}{5} \end{pmatrix} + \ln\frac{3}{8} = \frac{18}{5}x_1 - \frac{6}{5}x_2$

$-\frac{24}{5} + \ln\frac{3}{8}$

Разделяющая поверхность:

$\delta_0(X) = \delta_1(X)$

$\frac{8}{5}x_1 - \frac{6}{5}x_2 - \frac{4}{5} + \ln\frac{5}{8} = \frac{18}{5}x_1 - \frac{6}{5}x_2 - \frac{24}{5} + \ln\frac{3}{8}$

$-2x_1 + 4 + \ln\frac{5}{3} = 0$

$x_1 = 2 + \frac{\ln\left(\frac{5}{3}\right)}{2} \quad (1)$

Квадратичные дискриминантные функции:

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

$\ln \hat{Pr}\{Y=0\}$

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

Обратные матрицы:

$$\hat{\Sigma}_0^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$$

$$\left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 \end{array}\right) \xrightarrow{(2)-\frac{1}{2}(1)} \left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{2} & 1 \end{array}\right) \xrightarrow{(2)\cdot 4}$$

$$\left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & 1 & -2 & 4 \end{array}\right) \xrightarrow{(1)-\frac{1}{2}(2)} \left(\begin{array}{cc|cc} 1 & 0 & 2 & -2 \\ 0 & 1 & -2 & 4 \end{array}\right)$$

$$\hat{\Sigma}_1^{-1} = \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix}$$

$$\left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 & 1 \end{array}\right) \xrightarrow{(2)-\frac{1}{2}(1)} \left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & \frac{3}{4} & -\frac{1}{2} & 1 \end{array}\right) \xrightarrow{(2)\cdot\frac{4}{3}} \left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{4}{3} \end{array}\right)$$

$$\xrightarrow{(1)-\frac{1}{2}(2)} \left(\begin{array}{cc|cc} 1 & 0 & 4/3 & -2/3 \\ 0 & 1 & -2/3 & 4/3 \end{array}\right)$$

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \frac{8}{5} & -\frac{6}{5} \\ -\frac{6}{5} & \frac{12}{5} \end{pmatrix}$$

$$\left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{2}{3} & 0 & 1 \end{array}\right) \xrightarrow{(2)-\frac{1}{2}(1)} \left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & \frac{5}{12} & -\frac{1}{2} & 1 \end{array}\right) \xrightarrow{(2)\cdot\frac{12}{5}} \left(\begin{array}{cc|cc} 1 & \frac{1}{2} & 1 & 0 \\ 0 & 1 & -\frac{6}{5} & \frac{12}{5} \end{array}\right)$$

$$\xrightarrow{(1)-\frac{1}{2}(2)} \left(\begin{array}{cc|cc} 1 & 0 & \frac{8}{5} & -\frac{6}{5} \\ 0 & 1 & -\frac{6}{5} & \frac{12}{5} \end{array}\right)$$

№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) методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравн-е разделяющей пов-ти,
2) методом кв-го дискриминантного анализа построить дискриминантные функции. Изобразить точки и разделяющие пов-ти (кривые)

Вероятности классов: $\widehat{Pr}\{Y=0\} = \frac{5}{8}$

$\widehat{Pr}\{Y=1\} = \frac{3}{8}$

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

$\hat{\mu}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

Выборочные матрицы ковариации для каждого класса:

$$\hat{\Sigma}_0 = \frac{1}{N_0 - 1} \sum_{y^{(i)}=0} (x^{(i)} - \hat{\mu}_0)(x^{(i)} - \hat{\mu}_0)^T = \frac{1}{4}\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

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

Матрица ковариации:

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