Задание 9 Аракчеев Сергей

Дано

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

$N=8 \quad K=2$

$N_0=5 \quad N_1=3$

1) Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности

Вероятности классов

$$\widehat{Pr}\{Y=0\}=\frac{5}{8} \qquad \widehat{Pr}\{Y=1\}=\frac{3}{8}$$

Средние для классов

$$\hat{\mu}_0=\begin{pmatrix}\frac{1+2+2}{5}\\ \frac{-1+1}{5}\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$$

$$\hat{\mu}_1=\begin{pmatrix}\frac{2+4+3}{3}\\ \frac{1+2}{3}\end{pmatrix}=\begin{pmatrix}3\\1\end{pmatrix}$$

Выборочные матрицы ковариации для каждого класса

$$\hat{\Sigma}_0=\frac{1}{N_0-1}\sum_{y^{(i)}=0}\left(x^{(i)}-\hat{\mu}_0\right)\left(x^{(i)}-\hat{\mu}_0\right)^T=\frac{1}{4}\left(\begin{pmatrix}-1\\-1\end{pmatrix}\begin{pmatrix}-1&-1\end{pmatrix}+\begin{pmatrix}-1\\0\end{pmatrix}\begin{pmatrix}-1&0\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1&0\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1&1\end{pmatrix}\right)=\frac{1}{4}\left(\begin{pmatrix}1&1\\1&1\end{pmatrix}+\begin{pmatrix}1&0\\0&0\end{pmatrix}+\begin{pmatrix}1&0\\0&0\end{pmatrix}+\begin{pmatrix}1&1\\1&1\end{pmatrix}\right)=\frac{1}{4}\begin{pmatrix}4&2\\2&2\end{pmatrix}=\begin{pmatrix}1&0{,}5\\0{,}5&0{,}5\end{pmatrix}$$

$$\hat{\Sigma}_1=\frac{1}{N_1-1}\sum_{y^{(i)}=1}\left(x^{(i)}-\hat{\mu}_1\right)\left(x^{(i)}-\hat{\mu}_2\right)^T=\frac{1}{2}\left(\begin{pmatrix}1&1\\1&1\end{pmatrix}+\begin{pmatrix}1&0\\0&0\end{pmatrix}+\begin{pmatrix}0&0\\0&1\end{pmatrix}\right)=\frac{1}{2}\begin{pmatrix}2&1\\1&2\end{pmatrix}=\begin{pmatrix}1&\frac{1}{2}\\\frac{1}{2}&1\end{pmatrix}$$

Оцениваем матрицу ковариации

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

1) $\left(\begin{array}{cc|cc} 1 & 0,5 & 1 & 0 \\ 0,5 & 0,5 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 0,5 & 1 & 0 \\ 1 & 1 & 0 & 2 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 0,5 & 1 & 0 \\ 0 & 0,5 & -1 & 2 \end{array}\right) \Rightarrow$

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

2) $\left(\begin{array}{cc|cc} 1 & 0,5 & 1 & 0 \\ 0,5 & 0,5 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 0,5 & 1 & 0 \\ 0 & 0,75 & -0,5 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{4}{3} & -\frac{2}{3} \\ 0 & 0,75 & -\frac{1}{2} & 1 \end{array}\right) \Rightarrow$

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

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

$\Rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 1,6 & -1,2 \\ 0 & 1 & -1,2 & 2,4 \end{array}\right)$

$$\hat\Sigma_0^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}, \quad \hat\Sigma_1 = \frac{1}{3}\begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}, \quad \hat\Sigma_2 = \frac{1}{5}\begin{pmatrix} 8 & -6 \\ -6 & 12 \end{pmatrix}$$

Линейные дискриминантные функции

$$\delta_0(x) = x^T \hat\Sigma \hat\mu_0 - \frac{1}{2}\hat\mu_0 + \hat\Sigma^{-2}\hat\mu_0 + \ln \hat{Pr}\{Y=0\} =$$

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

$$= \frac{1}{5}(x_1 x_2)\begin{pmatrix} 8 \\ -6 \end{pmatrix} - \frac{1}{10}(1\ 0)\begin{pmatrix} 8 \\ -6 \end{pmatrix} + \ln\frac{5}{8} = \frac{1}{5}(8x_1 - 6x_2) - \frac{8}{10}$$

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

$$\delta_1(x) = x^T \hat\Sigma \hat\mu_1 - \frac{1}{2}\hat\mu_1 + \hat\Sigma\hat\mu_1 + \ln\hat{Pr}\{Y=1\} =$$

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

$= \frac{1}{5}(x_1\ x_2)\begin{pmatrix}18\\-6\end{pmatrix} - \frac{1}{16}(3\ \ 1)\begin{pmatrix}18\\6\end{pmatrix} + \ln\frac{3}{8} = \frac{1}{5}(18x - 6x_2)$

$- \frac{1}{10}\cdot 48 + \ln\frac{3}{8} = \frac{18}{5}x_1 - \frac{6}{5}x_2 - \frac{24}{5} + \ln\frac{3}{8}$

Разделяющая поверхность - прямая $\delta_0(x) = \delta_1(x)$

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

$\Rightarrow \frac{10}{5}x_1 - \frac{20}{5} + \ln\frac{5}{8} - \ln\frac{3}{8} = 0 \Rightarrow 2x_1 - 4 + \ln\frac{5}{3}$

Квадратичные дискриминантные функции

$\delta_0(x) = -\frac{1}{2}\ln\det\hat\Sigma_0 - \frac{1}{2}(x-\hat\mu_0)^T\hat\Sigma_0^{-1}(x-\hat\mu_0) + \ln Pr(Y=0)$

$\Rightarrow \det\hat\Sigma_0 = \begin{vmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{vmatrix} = 0{,}5 - 0{,}25 = \frac{1}{4}$

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

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

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

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

$\delta(x) = -\frac{1}{2}\ln\det\hat\Sigma - \frac{1}{2}(x-\hat\mu_1)^T\hat\Sigma_1^{-1}(x-\hat\mu_1) + \ln Pr$

$\{Y=1\} = -\frac{1}{2}\ln\frac{3}{4} - \frac{1}{2}(x-3\ \ x_2-1)\frac{1}{3}\begin{pmatrix}4 & -2\\ -2 & 4\end{pmatrix}\begin{pmatrix}x_1-3\\ x_2-1\end{pmatrix} + \ln\frac{3}{8} =$

$= -\frac{1}{2}\ln\frac{3}{4}(x-3\ \ x_2-1)\begin{pmatrix}2x-x_2-5\\ -x_1+2x_2+1\end{pmatrix} + \ln\frac{3}{8} = \ln\frac{3}{8} + \ln\frac{2}{\sqrt3}$

$- \frac{1}{3}\left((x-\frac{3}{2})(2x-x_2-5) + (x_2-1)(-x+2x_2+1)\right) - \frac{1}{2}\ln\frac{3}{16} -$

$- \frac{1}{3}(2x_1^2 - x_1x_2 - 5x_1 - 6x_1 + 3x_2 + 15 - x_1x_2 + 2x_2^2 + x_2 + x_1 +$

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

$+ \frac{1}{2}\ln\frac{3}{16}$

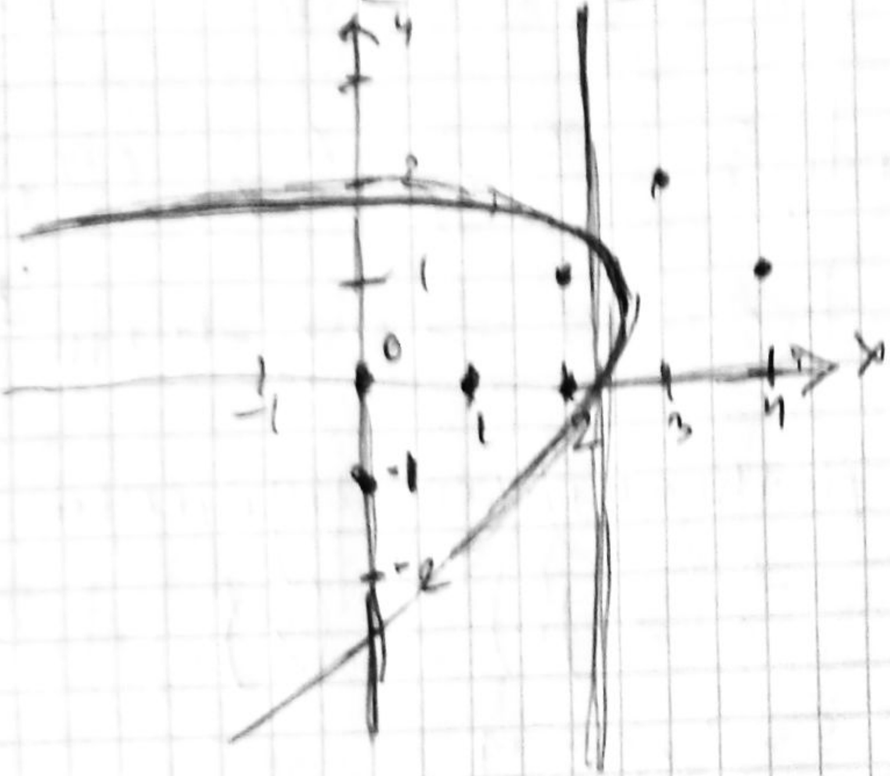
Разделяющая поверх - на парабола $\delta_0 x - \delta_1(x)$

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

$+ \frac{10}{3}x_1 - \frac{2}{3}x_2 + \frac{2}{3}x_1x_2 - \frac{14}{3} + \frac{1}{2}\ln\frac{3}{16}$

$$\frac{1}{3}\left(x_1^2+4x_2^2+4x+4x_2-4x_1x_2-11\right)+\frac{1}{2}\ln\frac{3}{25}=0 \Rightarrow$$

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



Аракчеев Сергей

Задание 16

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

$$X = \begin{pmatrix} 4 & 3 \\ 0 & 1 \\ -2 & -3 \\ 2 & -1 \end{pmatrix}$$

Найти главные направления и дисперсии по главным компонентам.

Решение:

Найдём $\bar{x}$ вектор, содержащий ср. знач. столбцов матрицы

$$\bar{x} = ((4+0-2+2)/4, (3+1-3-1)/4) = (1, 0)$$

Вычислим центрированную матрицу $X_c = X - M$, где $M = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$

$$X_c = \begin{pmatrix} 4 & 3 \\ 0 & 1 \\ -2 & -3 \\ 2 & -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ -1 & 1 \\ -3 & -3 \\ 1 & -1 \end{pmatrix}$$

$$C = X_c^T X_c = \begin{pmatrix} 3 & -1 & -3 & 1 \\ 3 & 1 & -3 & -1 \end{pmatrix} \begin{pmatrix} 3 & 3 \\ -1 & 1 \\ -3 & -3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 20 & 16 \\ 16 & 20 \end{pmatrix}$$

$$\det(C - \lambda I) = \begin{pmatrix} 20-\lambda & 16 \\ 16 & 20-\lambda \end{pmatrix} = (20-\lambda)^2 - 16^2 =$$

$$= 400 - 40\lambda + \lambda^2 - 256 = \lambda^2 - 40\lambda + 144 = (\lambda - 4)(\lambda - 36) = 0$$

Собственные числа: $\lambda_1 = 4$, $\lambda_2 = 36$

1) $\lambda_1 = 4$

$$\begin{pmatrix} 20-4 & 16 \\ 16 & 20-4 \end{pmatrix} = \begin{pmatrix} 16 & 16 \\ 16 & 16 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow$$

$$\Rightarrow x_1 = -x_2 \Rightarrow V_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

2) $\lambda_2 = 36$

$$\begin{pmatrix} 20-36 & 16 \\ 16 & 20-36 \end{pmatrix} = \begin{pmatrix} -16 & 16 \\ 16 & -16 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow V_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Нормируем собственные вектора

$$\overline{V_1} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix} \qquad \overline{V_2} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

Найдём дисперсии по главным компонентам, $N=4$

$$\frac{1}{N-1}\lambda_1 = \frac{4}{3}$$

$$\frac{1}{N-1}\lambda_2 = \frac{36}{3} = 12$$

Задание 15

Дано:

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

С помощью наивного байесова классификатора оценить вероятности $Pr(Y=0 \mid X_1=1, X_2=1)$, $Pr(Y=1 \mid X_1=1, X_2=1)$

Оцениваем априорные вероятности:

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

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

Оцениваем условные вероятности:

$\widehat{Pr}\{X_1=0 \mid y=0\} = \frac{3}{5}$ $\qquad$ $\widehat{Pr}\{X_2=0 \mid Y=0\} = \frac{2}{5}$

$\widehat{Pr}\{X_1=1 \mid Y=0\} = \frac{2}{5}$ $\qquad$ $\widehat{Pr}\{X_2=1 \mid Y=0\} = \frac{3}{5}$

$\widehat{Pr}\{X_1=0 \mid Y=1\} = \frac{2}{5}$ $\qquad$ $\widehat{Pr}\{X_2=0 \mid Y=1\} = 0$

$\widehat{Pr}\{X_1=1 \mid Y=1\} = \frac{3}{5}$ $\qquad$ $\widehat{Pr}\{X_2=1 \mid Y=1\} = 1$

Используя основное предположение наивного байесова классификатора, получаем:

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

$$= \frac{Pr\{X_1=1, Y=0\}\ Pr\{X_2=1, Y=0\}\ Pr\{Y_1=0\}}{Pr\{X_1=1, X_2=1\}} =$$

$$= \frac{\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{\frac{3}{25}}{\frac{21}{50}} = \frac{3 \cdot 50}{21 \cdot 25} = \frac{2}{7}$$

$$Pr\{Y=1 \mid X_1=1, X=1\} =$$

$$= \frac{Pr\{X_1=1, Y=1\} \cdot Pr\{X_2=1, Y=1\}\ Pr\{Y_1=1\}}{Pr\{X_1=1, X_2=1\}} =$$

$$= \frac{\frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{25} + \frac{3}{10}} = \frac{\frac{3}{10}}{\frac{21}{50}} = \frac{3 \cdot 50}{21 \cdot 10} = \frac{5}{7}$$

Задание 19

Пусть в задаче классификации на K классов $\{1, 2 \ldots K\}$ последний слой нейронной сети вычисляет softmax-функцию

$$g_k(s_1, s_2 \ldots s_K) = e^{s_k} \Big/ \sum_{l=1}^{K} e^{s_l}$$

В качестве штрафа используется кросс-энтропия

$$R^{(i)} = -\sum_{k=1}^{K} I(y^{(i)} = k) \ln g_k(s_1, s_2 \ldots s_K) \text{ — штраф,}$$

где $g_k(s_1, s_2 \ldots s_K)$ — softmax ф-ция

Доказать, что:

1) $$\frac{\partial g_k}{\partial s_l} = g_k\left(I(k=l) - g_l\right)$$

$\exists\, k \neq l$:

$$\frac{\partial g_k}{\partial s_l} = \frac{0 - e^{s_k} \cdot e^{s_l}}{\left(\sum e^{s_i}\right)^2} = -g_k \cdot g_l$$

$\exists\, k = l$:

$$\frac{\partial g_k}{\partial s_k} = \frac{e^{s_k} \cdot \sum e^{s_i} - e^{s_k} \cdot e^{s_k}}{\left(\sum e^{s_i}\right)^2} = g_k - g_k^2 =$$

$$= g_k(1 - g_k) \Rightarrow$$

$$\Rightarrow \frac{\partial g_k}{\partial s_l} = g_k\left(I(k=l - g_l)\right)$$

2) $$\frac{\partial R^{(i)}}{\partial g_k} = \frac{\sum^{K} I(y^{i} = l)}{g_k} = \frac{-I(y^{(i)})}{g_k},$$ ввиду того, что только одно значение будет равно 1, упрощаем выражение

3) $$\frac{\partial R^{(i)}}{\partial s_l} = g_l - I(l = y^{(i)})$$

$$\frac{\partial k^{(i)}}{\partial s1} = \frac{-I(y^{(i)}=1)}{g1} \cdot \frac{\partial g1}{\partial s1} - \sum_{j\neq 1} \frac{I(y^{(i)}=j)}{gj} \cdot$$

$$\cdot \frac{\partial gj}{\partial s1} = \frac{-I(y^{(i)}=1)}{gj} \cdot \frac{\cancel{\partial g1}}{\cancel{\partial s1}} g1(1-g1) + \sum_{j\neq 1}$$

$$\frac{I(y^{(i)}-j)}{gj} \cdot g1 \cdot gj = -I(y^{(i)}=1) + g1 - I(y^{(i)}=1) + g1 \cdot$$

$$\cdot \sum_{j\neq 1} I(y^{(i)}=j) = -I(y^{(i)}=1) + g1 \cdot \sum I(y^{(i)}=j) =$$

$$= g1 - I(y^{(i)}=1)$$