Машинное обучение

№35

Δ Задача восстановления регрессии

Функция потерь $L(y', y) = (y' - y)^2$

E - мат ожидание.

фактический выход   истинный выход

Д! $f^*(x) = \arg\min_c E((Y-c)^2 \mid \underbrace{X = x}_{\text{будем опускать}})$, то

$f^*(x) = E(Y \mid X = x)$

$E(Y-c)^2 = E(Y^2) - 2cE(Y) + c^2$

$(E(Y-c)^2)'_c = -2E(Y-c) = 2(c - E(Y)) = 0$

Выразим $c = E(Y) \Rightarrow f^*(x) = E(Y \mid X = x)$

и получили $\arg\min_c$

$Q(f^*) = \int E((\underset{E(Y)}{f^*(x)} - Y)^2 \mid x) p(x) dx =$

$= \int E(Y)^2 - 2E(Y)^2 + E(Y^2) \mid X) p(x) dx$ $\quad = E(E(Y)^2 - 2E(Y)Y + Y^2)$

$= \int \cancel{\ldots} + D_Y(Y|x) p(x) dx = D_Y(Y|x) E(x)$

№36

$L(y', y) = [y' - y]$ $\quad f(x) = \text{median}(Y \mid X = x)$

доставляет минимум среднему риску

$R(f) = \iint |f(x) - y| \, p(y|x) p(x) \, dy \, dx =$

$= \int E(|f(x) - Y| \mid x) p(x) dx$

Пусть $f(x) = c$ что $f^*(x) = \arg\min_c E(|c - Y| \mid x)$ Пусть $M = c$

Пусть $M = \text{median}(Y \mid X = x)$

$\Delta \; E(|c - Y|) = \int_{-\infty}^{+\infty} |c - Y| p(y) dy = \int_{-\infty}^{c} (c - Y) p(y) dy - \int_{c}^{+\infty} (c - Y) p(y) dy$

$= \int_{-\infty}^{M} (c - Y) p(y) dy - \int_{c}^{M} |c - Y| p(y) dy - \int_{c}^{M} |c - Y| p(y) dy + \int_{M}^{+\infty} |c - Y| p(y) dy$

$$= \int_{-\infty}^{M} (c + M - M - Y)\, p(y)\, dy \pm \cancel{\int_c^{M} + \int_M^{+\infty}} - 2\int_c^{M} (c-Y)\, p(y)\, dy -$$

$$= \int_{-\infty}^{M} |M - Y|\, p(y)\, dy + \int_{-\infty}^{M} |c-M|\, p(y)\, dy \qquad \boxed{- \int_M^{+\infty} (c + M - M - Y)\, p(y)\, dy}$$

$$+ \int_c^{M} |M-Y|\, p(y)\, dy + \int_c^{M} |c-M|\, p(y)\, dy - \int_c^{M} |c-Y|\, p(y)\, dy$$

$$+ \int_M^{+\infty} |M-Y|\, p(y)\, dy + \int_M^{+\infty} |c-M|\, p(y)\, dy$$

$$= \underbrace{E(|M-Y|)}_{\text{Не зависит от } c} + \underbrace{E(|M-c|)}_{\text{при } c=M \text{ получаем } 0.} - \cancel{\int (c-Y)\, p(y)\, dy}$$

$$= \int_{-\infty}^{M} (c-M)\, p(y)\, dy + \int_{-\infty}^{M} (M-Y)\, p(y)\, dy - 2\int_c^{M} (c-Y)\, p(y)\, dy$$

$$- \int_M^{+\infty} (c-M)\, p(y)\, dy + \int_M^{+\infty} (M-Y)\, p(y)\, dy =$$

$\downarrow E(M-Y)$

$$= E(M-Y) - 2\int_c^{M} (c-Y)\, p(y)\, dy + \underbrace{\int_{-\infty}^{M} (c-M)\, p(y)\, dy - \int_M^{+\infty} (c-M)\, p(y)\, dy}_{(c-M)\cdot\frac{1}{2} - \frac{1}{2}(c-M)}$$

$$= \underbrace{E(M-Y)}_{\text{не зависит от } c} - \underbrace{2\int_c^{M} (c-Y)\, p(y)\, dy}_{+2\int_c^{M} (Y-c)\, p(y)\, dy \ \geq 0}$$

$= 0$, если $c = M$

ч.т.д.

№37.

Найти функцию потерь, чтобы $\arg\min R(f)$ = условная мода

Пусть $L(y', y) = H(y', y) = \begin{cases} 0 & y' = y \\ 1 & y' \neq y \end{cases}$

$$R(f) = \iint H(f, Y)\, p(y|x)\, dy\, p(x)\, dx = \int E(\underbrace{H(f, Y)}_{\sim 0} | x)\, p(x)\, dx$$

$$f^*(x) = \arg\min_c E(H(c, Y)|x)$$

$$E(H(c, Y)) = 0 \cdot P_r(c = Y) + 1 \cdot P_r(c \neq Y)$$

$$\Delta\ \arg\min_c P_r(c \neq Y) \Rightarrow c = \mathrm{mode}(Y|x) = f^*(x)$$

чтобы вероятность была минимальной нужно чтобы с встречалось чаще

№38

Два класса $\{0, 1\}$ $L(y', y)$: $L(0,0) = L(1,1) = 0$, $L(1,0) = l_0$ $L(0,1) = l_1$

Д! Байесов классификатор $f^*(x)$ удв. условию

$$f(x) = \arg\max_{y \in \{0,1\}} l_y P_r(y|x)$$

$$\Delta\ R(f) = \int \Big(\sum_{y=0}^{1} L(f(x), y) \cdot P_r(y|x) \Big)\, p(x)\, dx$$

$$f^*(x) = \arg\min_c \Big(L(\underset{c}{f(x)}, 0) \cdot P_r(y|x) + L(\underset{c}{f(x)}, 1) \cdot P_r(y|x) \Big)$$

Нужно, чтобы сумма достигала минимума $\Rightarrow$

$\Rightarrow$ нужно занулить $L(f(x), y)$ перед максимальным $P_r(y|x)$

$$\Rightarrow f^*(x) = \arg\max_{y \in \{0,1\}} l_y P_r(y|x)$$

№38

Найти $f^*(x)$ если $h(y';y) = l_{y'y}$ $(y';y = 1,2..K)$

$$R(f) = \int \left[\sum_{y=1}^{K} h(f(x),y)\Pr(y|x)\right] p(x)\,dx$$

$$f^*(x) = \underset{f(x)}{\arg\min} \sum_{y=1}^{K} h(f(x),y)\Pr(y|x) = \arg\min \sum_{y=1}^{K} l_{yy'}\Pr(y|x)$$

Попав в один из классов $l_{yy'}=0$ => нужно минимизировать оставшуюся сумму

$$f^*(x) = \underset{y'}{\arg\min} \sum_{\substack{y\in Y\\ y\neq y'}} l_{yy'}\Pr(y|x)$$

№1.

$$\frac{\partial g}{\partial x} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & & & \cdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \qquad g = \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$$

1. Д!, $a \in \mathbb{R}^n$ $x \in \mathbb{R}^n$, то $\frac{\partial (a^T x)}{\partial x} = a$

$a^T x = a_1 x_1 + a_2 x_2 + \ldots a_n x_n$

$$\frac{\partial a^T x}{\partial x} = \frac{\partial}{\partial x}(a_1 x_1 + a_2 x_2 \ldots a_n x_n) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a$$

2. $A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \cdots & & \cdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$ $x \in \mathbb{R}^n$ $\qquad \frac{\partial Ax}{\partial x} = A$

$$Ax = (a_{ij})(x_{ij}) = \begin{pmatrix} a_{11} x_1 + a_{12} x_2 + \ldots a_{1n} x_n \\ \cdots \\ a_{m1} x_1 + \ldots a_{mn} x_n \end{pmatrix}$$

$$\frac{\partial Ax}{\partial x} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \cdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$$

~~3. $A = \begin{pmatrix} a_{11} \ldots a_{1n} \\ a_{n1} \ldots a_{nn} \end{pmatrix}$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\frac{\partial (x^T A x)}{\partial x} = (A + A^T) x$~~

~~$x^T A = \begin{pmatrix} x_1 a_{11} + \ldots + x_n a_{n1} \\ \ldots \\ x_1 a_{1n} + \ldots + x_n a_{nn} \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_n \end{pmatrix} \to (x_1^2 a_{11} + \ldots + x_n x_1 a_{n1})$~~

~~$(x_1 a_{11} + \ldots x_n a_{n1}, \ldots x_1 a_{1n} + \ldots + x_n a_{nn}) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1^2 a_{11} + \ldots x_n x_1 a_{n1} \\ \ldots \\ x_1 x_n a_{1n} + \ldots + x_n^2 a_{nn} \end{pmatrix}$~~

~~$\frac{\partial (x^T A x)}{\partial x} = 2 x_1 a_{11} + \ldots x_n a_{n1}, \; x_1 a_{12}, \; x_1 a_{13} \ldots x_1 a_{n1}$~~

~~$a_{21} x_1 + 2 x_2 a_{22} + \ldots x_n a_{n1}$~~

~~$Ax = \sum_{j=1}^{n} a_{1j} x_j \ldots \sum_{j=1}^{n} a_{nj} x_j$ $\quad (x_1 x_2 \ldots x_n) \begin{pmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{nj} x_j \end{pmatrix} = \begin{pmatrix} x_1 \cdot \sum a_{1j} x_j \\ x_2 \cdot \sum a_{2j} x_j \\ \vdots \\ x_n \sum a_{nj} x_j \end{pmatrix}$~~

$x^T A x = \sum_{i=1}^{n} x_i \cdot \sum_{j=1}^{n} a_{ij} x_j \qquad \frac{\partial x^T A x}{\partial x} =$

3) $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$ $\quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\quad A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ \vdots & & & \vdots \\ a_{1n} & \dots & & a_{nn} \end{pmatrix}$

$$x^T A x = (x_1\ x_2 \dots x_n) \cdot \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + \dots a_{nn}x_n \end{pmatrix} =$$

$$= x_1 \cdot \sum_{j=1}^{n} a_{1j}x_j + \dots x_n \cdot \sum_{j=1}^{n} a_{nj}x_j$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij}x_j$$

$$\frac{\partial(x^T A x)}{\partial x} = \begin{pmatrix} 2a_{11}x_1 + \dots a_{1n}x_n + a_{21}x_2 + a_{31}x_3 + \dots a_{n1}x_n \\ \vdots \\ x_1 a_{1n} + x_2 a_{2n} + \dots + 2a_{nn}x_n \end{pmatrix}$$

$$\frac{\partial g}{\partial x} = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix}$$

$$= (A + A^T)\, x$$

4) $x \in \mathbb{R}^n$, $\frac{\partial \|x\|^2}{\partial x} = 2x$ D!

$\|x\| = \sum_{i=1}^{n} x_i^2$

$$\frac{\partial \sum_{i=1}^{n} x_i^2}{\partial x} = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} = 2x$$

5) $$\frac{\partial g(x)}{\partial (x)} = \begin{pmatrix} g'_{x_1} & 0 & \dots & 0 \\ 0 & g'_{x_2} & \dots & \\ 0 & \dots & & g'_{x_n} \end{pmatrix}$$

$$g(x) = \begin{pmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{pmatrix}$$

$$\frac{\partial g}{\partial x} = \begin{pmatrix} \frac{\partial g(x_1)}{\partial x_1} & \overset{0}{\frac{\partial g(x_1)}{\partial x_2}} & \dots & \overset{0}{\frac{\partial g(x_n)}{\partial x_n}} \\ \underset{0}{\frac{\partial g(x_n)}{\partial x_1}} & \dots & & \frac{\partial g(x_n)}{\partial x_n} \end{pmatrix}$$

6) $h: \mathbb{R}^n \to \mathbb{R}^m$
$g: \mathbb{R}^m \to \mathbb{R}^p$
$x \in \mathbb{R}^n$

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

$$\cancel{h(x) = \begin{pmatrix} h(x_1) \\ h(x_n) \end{pmatrix}} \qquad \cancel{g(h(x)) = \begin{pmatrix} g(h(x_1)) \\ g(h(x_2)) \\ g(h(x_n)) \end{pmatrix}}$$

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

$$\frac{\partial g(h(x))}{\partial x} = \begin{pmatrix} \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_n} \\ \dots & \dots & \dots \\ \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_n} \end{pmatrix} = \frac{\partial g(h(x))}{\partial h} \cdot \frac{\partial h(x)}{\partial x}$$

№16

x_1	4	0	−2	2
x_2	3	1	−3	−1

Главные напр.
Дисперсия

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

1. Центрируем данные

$$\bar{X} = \frac{1}{N}\left(\sum_1^4 x_{1i}, \sum_1^4 x_{2i}\right) = \frac{1}{4}(4, 0)$$

$$\bar{X} = (1, 0)$$

$$X_c = X - (\bar{X}) = \begin{pmatrix} 3 & 3 \\ -1 & 1 \\ -3 & -3 \\ 1 & -1 \end{pmatrix}$$

2. Найдем

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

Выборочная матрица ковариации $\frac{1}{N-1} C = \frac{1}{3}\begin{pmatrix} 20 & 16 \\ 16 & 20 \end{pmatrix}$

3. Найдем собств числа и вектора матрицы C

$$\Delta = \begin{vmatrix} 20-\lambda & 16 \\ 16 & 20-\lambda \end{vmatrix} = 0 \qquad \Delta = \lambda^2 - 40\lambda + 144 = 0$$

$$\lambda_1 = 36 \quad \lambda_2 = 4$$

$36 = \lambda_1$: $\begin{pmatrix} -16 & 16 \\ 16 & -16 \end{pmatrix}\begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad \tilde{v}_1 = \tilde{v}_2 = 1 \quad v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$4 = \lambda_2$: $\begin{pmatrix} 16 & 16 \\ 16 & 16 \end{pmatrix}\begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

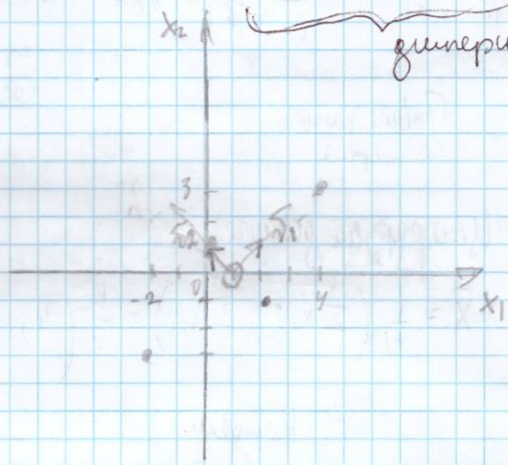
Нормировка: $\sqrt{c^2 + c^2} = 1 \quad \sqrt{2}\,c = 1 \quad c = \frac{\sqrt{2}}{2}$

$v_1 = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad v_2 = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 \\ +1 \end{pmatrix}$

v_1, v_2 – главные компоненты

$\frac{1}{N-1}\lambda_1 = \frac{1}{N-1}\sigma_1^2 = \frac{36}{3} = 12 \qquad \frac{1}{N-1}\lambda_2 = \frac{1}{N-1}\sigma_2^2 = \frac{4}{3}$

дисперсии по главным компонентам



№48

Влияние шума на качество предсказания

Пр-во признаков – одномерное

Обучающая выборка $x^{(0)} = 0$ $x^{(1)} = 1$

Добавляем шумовой признак $x^{(0)} = (0, \xi_1)$ $x^{(1)} = (1, \xi_2)$

Найти вероятность, что $x = (0.32, 0)$ ближе к $x^{(1)}$

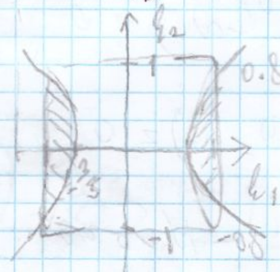
$$\rho(x^{(0)}, x)^2 = 0.32^2 + \xi_1^2$$

$$\rho(x^{(1)}, x)^2 = (1-0.32)^2 + \xi_2^2 = 0.68^2 + \xi_2^2$$

Найти $P\left(0.68^2 + \xi_2^2 \leq 0.32^2 + \xi_1^2\right)$

$$P\left(0.36 + \xi_2^2 \leq \xi_1^2\right) = \iint\limits_D f_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2$$

$D: 0.36 + \xi_2^2 \leq \xi_1^2$



$$= 4 \int_0^{0.8} d\xi_2 \int_{\sqrt{0.36+\xi_2^2}}^{1} \frac{1}{2} \cdot \frac{1}{2}\, d\xi_1$$

$$= \int_0^{0.8} 1 - \sqrt{0.36 + \xi_2^2}\, d\xi_2$$

$$\approx 0.20225$$

№4.1

так можно

Доказать, что $v_0 = \overline{x}$

Найти все v_0 доставляющих минимум сумме квадратов расстояний до искомого многообразия

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}$$

многообр. $W_k = v_0 + L(v_1, \dots v_k)$ $v_j \in \mathbb{R}^d$ $\|v_j\| = 1$ $v_j \perp v_{j'}$ $(j \neq j')$

где:
$$\sum_{i=1}^{N} dist^2(x^{(i)}, W_k) = \sum_{i=1}^{N} \|ort_{W_k} x^{(i)}\|^2 =$$
$$= \sum_{i=1}^{N} \|x^{(i)} - v_0 - \sum_{j=1}^{K} (x^{(i)} - v_0, v_j)\, v_j\|^2 \to min$$

v_j - векторы главных компонент можно находить итерационно

$$\underset{v_0}{argmin} \sum_{i=1}^{N} \|x^{(i)} - v_0\|^2$$

$$\frac{\partial\left(\sum_{i=1}^{N} \|x^{(i)} - v_0\|^2\right)}{\partial v_0} = -\sum_{i=1}^{N} 2(x^{(i)} - v_0) = 0$$
$$\Downarrow$$
$$\sum_{i=1}^{N} (x^{(i)} - v_0) = 0$$
$$\sum_{i=1}^{N} x^{(i)} - N v_0 = 0$$
$$v_0 = \frac{1}{N}\sum_{i=1}^{N} x^{(i)} = \overline{x}$$

чтд.

№4.2

Д! $X V_k = U_k \Sigma_k$

Известно $\underset{N\times d}{X} = \underset{N\times N}{U}\,\underset{N\times d}{\Sigma}\,\underset{d\times d}{V^T}$ $\quad V^T = V^{-1}$ $\quad V^T \cdot V_k = V^{-1} V_k = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & 1 \\ 0 & \dots & & \end{pmatrix}$

$$X V_k = U \Sigma V^T V_k \overset{?}{=} U_k \Sigma_k$$

$$\underset{N\times N}{U} \cdot \underset{N\times d}{\Sigma} \cdot \underset{d\times k}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \\ 0 & \dots & 1 \\ 0 & \dots & 0 \end{pmatrix}} = \underset{N\times N}{U} \cdot \underset{N\times k}{\begin{pmatrix} \Sigma_k \\ 0 \end{pmatrix}} = \underset{N\times N}{U} \cdot \underset{N\times k}{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & \dots & 0 \end{pmatrix}} \underset{k\times k}{\Sigma_k} = U_k \Sigma_k$$

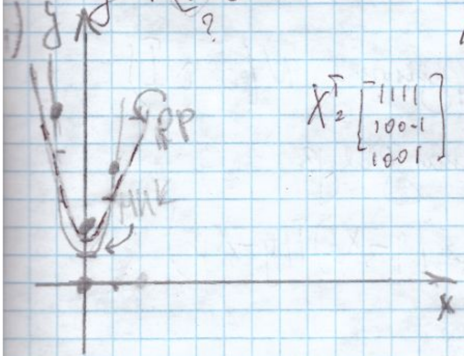
чтд.

№ 3.

Обучающая выборка:

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

?



2) МНК

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \qquad y = \begin{pmatrix} 4 \\ \cancel{4} \\ 0 \\ 2 \\ 6 \end{pmatrix}$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 12 \\ -2 \\ 10 \end{bmatrix}$$

Решим систему $X^T X \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{bmatrix} 12 \\ -2 \\ 10 \end{bmatrix} = X^T y$

$$\begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{bmatrix} 12 \\ -2 \\ 10 \end{bmatrix} \qquad \beta = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$$

$$f(x) = 1 - x + 4x^2$$

3) Ридж-регрессия с $\lambda = 1$

Регуляризованная система: $(X^T X + \lambda I)\beta = X^T y$

$$\begin{bmatrix} 4+\lambda & 0 & 2 \\ 0 & 2+\lambda & 0 \\ 2 & 0 & 2+\lambda \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 12 \\ -2 \\ 10 \end{pmatrix}$$

$$\begin{bmatrix} 5 & 0 & 2 \\ 0 & 3 & 0 \\ 2 & 0 & 3 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{bmatrix} 12 \\ -2 \\ 10 \end{bmatrix} \qquad \beta \approx \begin{bmatrix} 1.45 \\ -0.7 \\ 2.36 \end{bmatrix}$$

$$f(x) = 1.45 - 0.7x + 2.36x^2$$

№4.

Δ Задача восстановление регрессии

y распределен по нормальному закону:

$$N(X\beta, \sigma^2 I)$$

где β имеет априорное распределение $N(0, \tau I)$

Найти апостериорное распределение для β

Имеем $P_{y,X|\beta}(y,X|\beta) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \cdot e^{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)}$

$$P_\beta(\beta) = \frac{1}{(\sqrt{2\pi\tau})^d} e^{-\frac{1}{2\tau}\beta^T\beta}$$

$$P_{\beta|Y,X}(\beta|y) = \frac{P_{Y,X|\beta}(y,X|\beta)\cdot P_\beta(\beta)}{P_{Y,X}(y,X)} = \frac{\frac{1}{(\sqrt{2\pi})^{N+d}\sigma^N\tau^{\frac{d}{2}}} e^{-\frac{1}{2}\left(\frac{\|y-X\beta\|^2}{\sigma^2}+\frac{\|\beta\|^2}{\tau}\right)}}{P_{YX}(y,X)}$$

Δ $\frac{\|y-X\beta\|^2}{\sigma^2}+\frac{\|\beta\|^2}{\tau} = \frac{1}{\sigma^2}(y-X\beta)^T(y-X\beta)+\frac{1}{\tau}\beta^T\beta \ominus$

$\ominus \frac{1}{\sigma^2}(y^Ty - 2\beta^TX^Ty + \beta^TX^TX\omega) + \frac{\beta^T\beta}{\tau} = \frac{1}{\sigma^2}yy^T - \frac{2}{\sigma^2}\beta^TX^Ty$

$+\omega^T\underbrace{\left(\frac{1}{2\sigma^2}A^TA+\frac{1}{\tau}I\right)}_{\Sigma}\omega = \frac{1}{\sigma^2}y^Ty - \frac{2}{\sigma^2}\beta^TX^Ty+\omega^T\Sigma\omega$

$\frac{1}{\sigma^2}\beta^TX^Ty = \beta^T\Sigma\mu \to \mu = \frac{1}{\sigma^2}\Sigma^{-1}X^Ty$

Получим $\underbrace{(\omega - \frac{1}{\sigma^2}\Sigma^{-1}X^Ty)^T\Sigma(\omega-\frac{1}{\sigma^2}\Sigma^{-1}X^Ty)} \sim N(\overset{\text{среднее}}{\mu}, \underset{\text{дисперсия}}{\Sigma^{-1}})$

$P(y,X|\beta) = const\cdot e(\omega-)$

Параметр регуляризации $\lambda = \frac{\sigma^2}{\tau}$

Δ $P_{(y,X|\beta)}$ - норм. распределение $\Rightarrow$ $Mode(\beta) = \beta^{ridge} = E(\beta)$

№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

Оценить $Pr(Y=0 \mid X_1=1, X_2=1)$

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

$$Pr(Y=0 \mid X_1=1, X_2=1) \Rightarrow \frac{\hat{Pr}(X_1=1 \mid Y=0)\,\hat{Pr}(X_2=1 \mid Y=0)\,\hat{Pr}(Y=0)}{\hat{Pr}(X_1=1, X_2=1)}$$

Условные вероятности:

$$\hat{Pr}(X_1=1 \mid Y=0) = \frac{2}{5}$$

$$\hat{Pr}(X_2=1 \mid Y=0) = \frac{3}{5}$$

$$\hat{Pr}(Y=0) = \frac{1}{2}$$

$$\hat{Pr}(X_1=1, X_2=1) = Pr(Y=0) \cdot Pr(X_1=1 \mid Y=0) \cdot Pr(X_2=1 \mid Y=0) + Pr(Y=1)\, Pr(X_1=1 \mid Y=1) \cdot Pr(X_2=1 \mid Y=1)$$

$$\hat{Pr}(X_1=1, X_2=1) = \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{3}{5} \cdot 1 = \frac{3}{25} + \frac{3}{10}$$

$$Pr(Y=0 \mid X_1=1, X_2=1) = \frac{\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{2}}{\frac{6+15}{50}} = \frac{2}{7}$$

$$Pr(Y=1 \mid X_1=1, X_2=1) = \frac{\hat{Pr}(X_1=1 \mid Y=1)\,\hat{Pr}(X_2=1 \mid Y=1) \cdot \hat{Pr}(Y=1)}{Pr(X_1=1, X_2=1)} =$$

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

№ 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

Построить дискриминантную функцию

Оценим вероятности классов: $\hat{P}_r(Y=0) = \frac{5}{8}$ $\hat{P}_r(Y=1) = \frac{3}{8}$

Оцениваем средние для классов при $Y=0$: $\hat{\mu}_0 = \begin{pmatrix} \frac{0+1+0+2+2}{5} \\ \frac{-1+0+0+0+1}{5} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$Y=1$ $\hat{\mu}_1 = \begin{pmatrix} \frac{2+4+3}{3} \\ \frac{0+1+2}{3} \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

Выборочные матрицы ковариации для каждого класса

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

$$= \frac{1}{4}\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

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

$$= \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Оцениваем матрицу ковариации

$$\hat{\Sigma} = \frac{1}{\underset{8\ \ \ 2}{N-K}} \sum_k \sum_{y^{(i)}=k} (x^{(i)} - \hat{\mu}_k)(x^{(i)} - \hat{\mu}_k)^T = \frac{1}{6}\left(\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right) =$$

$$= \frac{1}{6}\begin{pmatrix} 6 & 3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & \frac{2}{3} \end{pmatrix}$$

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

Линейные дискриминантные функции

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

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

$$+ \ln\left(\frac{3}{8}\right) = 1.6 x_1 - 1.2 x_2 - 0.8 - 0.47 =$$

$$= 1.6 x_1 - 1.2 x_2 - 1.27$$

$$\delta_1(x) = -\| - = 3.6 x_1 - 1.2 x_2 - 5.78$$

Разделяющая поверхность: $\delta_0(x) = \delta_1(x)$

$$2 x_1 + 0 \cdot x_2 - 4.51 = 0$$

$$x_1 = 2.255$$

Квадратичные дискриминантные функции

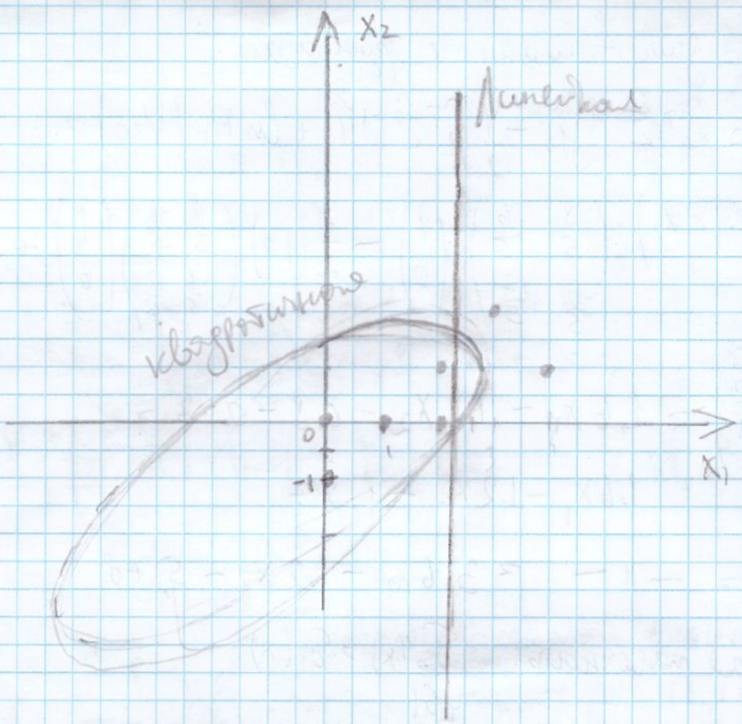
$$\delta_0(x) = -\frac{1}{2} \ln \Delta \hat{\Sigma}_0 - \frac{1}{2} (x - \hat{\mu}_0)^T \hat{\Sigma}_0^{-1} (x - \hat{\mu}_0) + \ln \hat{P_r}(Y=0)$$

$$= -x_1^2 + 2 x_1 x_2 + 2 x_1 - 2 x_2 - 2 x_2^2 - 0.78$$

$$\delta_1(x) = -\| - = -0.67 x_1^2 - 0.67 x_2^2 + 0.67 x_1 x_2 + 3.33 x_1 - 0.67 x_2 - 5.5$$

Разделяющая плоскость $\delta_0 = \delta_1$: $0.33 \cdot x_1^2 - 1.33 x_1 x_2 + 1.33 x_1 + 1.33 x_2 + 1.33 x_2^2 - 4.72 = 0 \quad | \cdot 3$

$$x_1^2 - 4 x_1 x_2 + 4 x_1 + 4 x_2 + 4 x_2^2 - 14 = 0$$



X_2
линейная
квадратичная
0
1
-1
X_1

№ 19.

Классификация на K классов

Есть функция $g_k(s_1, s_2 \ldots s_k) = \dfrac{e^{s_k}}{\sum\limits_{l=1}^{K} e^{s_l}}$

Штраф:

$$R^{(i)} = -\sum_{k=1}^{K} \mathbb{I}(y^{(i)}=k) \ln g_k(\ldots)$$

Д! 1) $\dfrac{\partial g_k}{\partial s_l} = g_k \cdot (\mathbb{I}(k=l) - g_l)$

$$\frac{\partial g_k}{\partial s_l} = k=l \quad \left(\frac{e^{s_l}}{\sum\limits_{l=1}^{K} e^{s_l}} \right)'_{s_l} = \frac{e^{s_l}}{(e^{s_1} + \ldots e^{s_l} + \ldots e^{s_k})}$$

$$= \frac{e^{s_l} \cdot \sum\limits_{i=1}^{K} e^{s_i} - e^{s_l} \cdot e^{s_l}}{\left(\sum\limits_{i=1}^{K} e^{s_i}\right)^2} = g_l - g_l^2 =$$

$$\overset{l=k}{=} g_k(1-g_l)$$

$k \neq l$

$$\left(\frac{e^{s_k}}{(e^{s_1} + \ldots e^{s_l} + \ldots e^{s_k})} \right)'_{s_l} = \frac{-e^{s_k} e^{s_l}}{\left(\sum\limits_{i=1}^{K} e^{s_i}\right)^2} =$$

$$= -g_k \cdot g_l = g_k(0 - g_l) \quad \text{ч.т.д.}$$

2) $\dfrac{\partial R^{(i)}}{\partial g_l} = -\dfrac{\mathbb{I}(y^{(i)})}{g_l}$

будет только один элемент с индексом l

$$\frac{\partial R^{(i)}}{\partial g_l} = \left(-\sum_{k=1}^{K} \mathbb{I}(y^{(i)}=k) \ln g_k(s_1 \ldots s_k) \right)'_{g_l} = -\frac{\mathbb{I}(y^{(i)}=k)}{g_l}$$

$$\left(-\mathbb{I}(y^{(i)}=k) \cdot \ln g_l \right)'_{g_l}$$

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

$$\frac{\partial R^{(i)}}{\partial s_l} = \left(-\sum_{k=1}^{K} I(y^{(i)}=k)\ln g_k(s_1, \dots s_l, s_k)\right)'_{s_l} =$$

$$= -\left[I(y^{(i)}=1)\cdot\frac{\frac{\partial g_1}{\partial s_l}}{g_1(\dots)} + \dots I(y^{(i)}=K)\cdot\frac{\frac{\partial g_K}{\partial s_l}}{g_K}\right] =$$

$$\frac{\partial g_k}{\partial s_l} = \begin{cases} -g_k g_l & k \neq l \\ g_k(1-g_k) & k = l \end{cases}$$

$$= -\left[I(y^{(i)}=1)\cdot\left(\frac{-g_1 g_l}{g_1}\right) + \dots + I(y^{(i)}=l)\cdot(1-g_l)\right] =$$

$$= g_l\left(\underbrace{\sum_{k=1}^{K} I(y^{(i)}=k)}_{=1}\right) + I(y^{(i)}=l)\cdot(\cancel{g_l}-1) =$$

$$= g_l - I(y^{(i)}=l)$$

№20

K-задач 2 классов предсказать $y_k \in \{0,1\}$ $k=1\dots K$

обучающая выборка $(x^{(i)}, y^{(i)})$

$$y^{(i)} = (y_1^{(i)}, y_2^{(i)} \dots y_K^{(i)}) \in \{0,1\}^K \quad i=1,2\dots N$$

$$g_k(s_k) = \frac{1}{1+e^{-s_k}} \quad k=1,2\dots K$$

Энтроп.: $$R^{(i)} = -\sum_{k=1}^{K} y_k^{(i)}\ln g_k + (1-y_k^{(i)})\ln(1-g_k)$$

Д! $$\frac{\partial R^{(i)}}{\partial s_l} = g_l - y_l^{(i)}$$

$$\frac{\partial R^{(i)}}{\partial s_l} = \left(\dots - y_l^{(i)}\ln g_l - (1-y_l^{(i)})\ln(1-g_l) - \dots\right)'_{s_l} =$$

$$= -y_l^{(i)}\cdot\frac{g_l'}{g_l} - (1-y_l^{(i)})\cdot\frac{(-1)\cdot g_l'}{1-g_l}$$

$$g_l' = \frac{e^{-s_l}}{(1+e^{-s_l})^2} = e^{-s_l}\cdot g_l^2 \qquad 1-g_l = \frac{1+e^{-s_l}-1}{1+e^{-s_l}} = \frac{e^{-s_l}}{1+e^{-s_l}}$$

$$\frac{\partial R^{(i)}}{\partial s_l} = +y_l^{(i)} e^{-s_l} g_l + (1-y_l^{(i)}) \frac{\cdot(-1)\cdot e^{-s_l} g_l^2}{1-g_l}$$

$$\frac{-e^{-s_l}}{1+e^{-s_l}} \qquad g_l^2 \cdot \frac{(1-g_l)\cdot g_l^{-1}}{(1-g_l)}$$

$$= -y_l^{(i)}(1-g_l) + (1-y_l^{(i)}) g_l$$

$$= -y_l^{(i)} + y_l^{(i)} g_l + g_l - y_l^{(i)} g_l = g_l - y_l^{(i)}$$

чтд

№21.

2 линейных слоя

Вычисление $\mathrm{softmax}(B(\sigma(Ax))) = g$

Функция потерь $\mathrm{logloss} = R(g)$

Потери на $x^{(i)}$: $\mathrm{logloss}(\mathrm{softmax}\, B(\sigma(Ax^{(i)})))$

Матричные формулы для backpropagation:

Пусть выход на первом слое $y = Ax$

после активации $z = \sigma(y)$

Выход на 2 слое: $t = Bz$

Применяем softmax: $g = \mathrm{softmax}(t)$

Нужно найти обновление весов:

Есть $\delta_x = \frac{\partial R^{(i)}}{\partial x}$ x - узлы сети

$$\frac{\partial R^{(i)}}{\partial x} = \frac{\partial R^{(i)}}{\partial g}\frac{\partial g}{\partial t}\cdot\frac{\partial t}{\partial z}\frac{\partial z}{\partial y}\frac{\partial y}{\partial x} = \frac{\partial R}{\partial g}\frac{\partial g}{\partial t}\cdot\frac{\partial (Bz)}{\partial z}\frac{\partial(\sigma(y))}{\partial y}\frac{\partial (Ax)}{\partial x}$$

$$= (g - y)\cdot B \cdot \mathrm{diag}(\sigma'(y))\cdot A$$

Обновление весов:

Весовые матрицы

$$\bar{A} = A - \alpha\frac{\partial R^{(i)}}{\partial A} \qquad \frac{\partial R^{(i)}}{\partial A} = \frac{\partial R^{(i)}}{\partial g}\frac{\partial g}{\partial t}\frac{\partial B(z)}{\partial z}\frac{\partial(\sigma(y))}{\partial y}\frac{\partial Ax}{\partial A} = \delta_y x$$

$$\bar{B} = B - \alpha\frac{\partial R^{(i)}}{\partial B} \qquad \frac{\partial R^{(i)}}{\partial B} = \frac{\partial R^{(i)}}{\partial g}\frac{\partial g}{\partial t}\frac{\partial Bz}{\partial B} = \delta_t z$$