KNN

$Y = \{1 \dots K\}$ — многокласс. класс-ия

$X = (x_i, y_i)_{i=1}^{\ell}$

Вводим ф-ию ρ → симметр. → метрику.

$\rho(u, x_u^{(1)}) \le \rho(u, x_u^{(2)}) \le \rho(u, x_u^{(2)}) \dots$
$\le \rho(u, x_u^{(n)})$

возвращаем

$$a(u) = \arg\max_{y' \in Y} \sum_{i=1}^{k} [y_u^{(i)} = y]$$

Какие проблемы

? $k = 5$

для $i = 1, 2$ $\quad \rho(u, x_u^{(i)}) \approx \underline{1}$.

$i = 3, 4, 5$ $\quad \underline{\rho(u, x_u^{(i)})} \approx 100$

$$a(u) = \arg\max_{y \in Y} \sum_{i=1}^{k} w_i [y_u^{(i)} = y]$$

1

$$w_i = \frac{k + 1 - i}{k}$$

положит.-ая, четная, гладкая

2

$$a(u) = \arg\max \sum_{i=1}^{k} \underbrace{K\left(\frac{\rho(u, x_u^{(i)})}{\underbrace{h}_{\text{ширина окна}}}\right)}_{w_i} [y_u^{(i)} = y]$$

$$\frac{\rho(u, x_u^{(i)})}{h} = r$$

пример

$$K(r) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} r^2\right\}$$

$\boxed{h}$ – гиперпараметр.

Ядровая регрессия

$$a(u) = \arg\min_{C \in R} \sum_{i=1}^{k} K(r) (C_u - y_u^{(i)})^2$$

$$\hat{C}_u = a(u)$$

$$2 \sum K(r) (C_u - y_u^{(i)}) = 0.$$

$$\sum K(r) C_u - \sum K(r) y_u = 0.$$

$$\hat{C}_u = \frac{\sum_{i=1}^{k} K(r) y_u^{(i)}}{\sum_{i=1}^{k} K(r)}$$

вектор, вектор

$$\rho(x, z) = \left(\sum_{j=1}^{d} |x_j - z_j|^2 \right)^{1/2}$$ ← евклидово расст.

метка класса

	y	$x^{(1)}$	$x^{(2)}$
1	1	0.1	$x_2^{(2)} \sim Unif[0;1]$ ← γ_1
2	0	0.5	$x_1^{(2)} \sim Unif[0;1]$ ← γ_2

y_{test}	x^1_{test}
1	0

$\rho_1(y_1, y_{test}) = (|0 - 0.1|^2)^{1/2} = 0.1$

$\rho_2(y_2, y_{test}) = 0.5$

$\rho_1 < \rho_2 \Rightarrow \underset{a(x^1_{test})}{\underset{\shortparallel}{y_{\cancel{test}}}} = 1$

(1) Посчитать вер-ств того, что с новым признаком $x^{(2)}$ объект y_2 окажется к y_{test} ближе.

$u = (0; 0)$, где 0 — это $x^{(1)}$, 0 — это $x^{(2)}$

$\rho_1 = 0.1^2 + \gamma_1^2$

$\rho_2 = 0.5^2 + \gamma_2^2$

$P(\rho_1 \geq \rho_2) =$

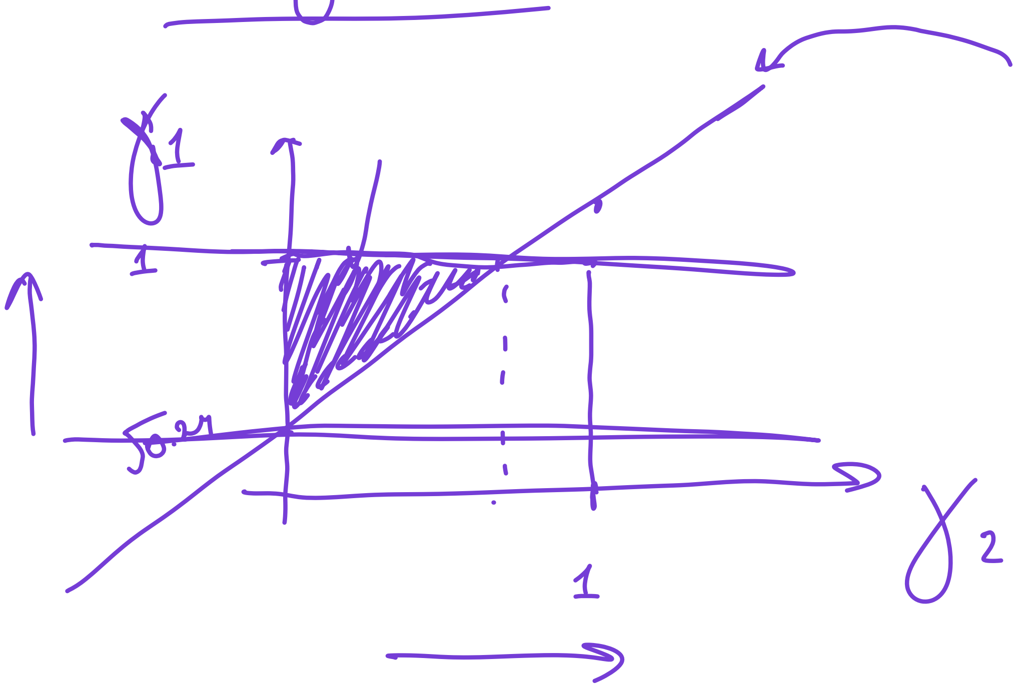
$$P(0.5^2 + Y_2^2 \le (0.1)^2 + Y_1^2) =$$

$$= P(Y_1^2 \ge 0.24 + Y_2^2) = P\left(Y_1 \ge \sqrt{0.24 + Y_2^2}\right)$$

$Y_1 \sim Unif[0;1]$

$Y_2 \sim Unif[0;1]$. $\quad Y_1 \perp\!\!\!\perp Y_2$

$\underline{P(Y_1 > 0.5)}$



$$Y_1 = \sqrt{0.24 + Y_2^2}$$

$$1 = \sqrt{0.24 + Y_2^2}$$

$$0.76$$

$$\Rightarrow Y_2 = \sqrt{0.76}$$

$$f_{Y_1 Y_2}(\cdot) = f_{Y_1} \cdot f_{Y_2} = 1$$

$$\int_0^{\sqrt{0.76}} \left(\int_{\sqrt{0.24 + x_2^2}}^{1} f_{Y_1 Y_2} \, dx_1 \right) dx_2 = \int_0^{\sqrt{0.76}} \left(1 - \sqrt{0.24 + x_2^2}\right) dx_2 \approx 0.27$$

Проклятие размерности

1NN

n - мерное пр-во

n - большое число

N - объектов, ∫ характеризуются n признаками

Они лежат в n-мерной сфере радиуса 1.

$\underline{a} = (0, 0, 0 \ldots)_n$

Вопрос: Какое расст. от точки a до её ближ. соседа?

1. Зависит от выборки
2. медианное расст-ие d_{med}

Какова вер-сть того, что все точки выборки окажутся от a дальше, чем d_{med}?

$$\frac{1}{2} = \left(\frac{k_n - k_n d^n}{k_n}\right)^N$$

с радиусом d

$$V_{\text{сферы}} = k_n \cdot r^n$$

$$V^{d}_{\text{сферы}} = k_n \cdot d^n$$

$$V^{1}_{\text{сферы}} = k_n \cdot 1^n$$

$$d = \Big(\underbrace{1 - \left(\tfrac{1}{2}\right)^{1/N}}_{0 < x < 1}\Big)^{1/n}$$

$$x^{1/n}$$

d - ? когда $n \to \infty$

$d \to \underline{1}$

<u>Метрики.</u>

Минковского

$$\rho_p(x, z) = \left(\sum_{j=1}^{d} |x_j - z_j|^p\right)^{1/p}$$

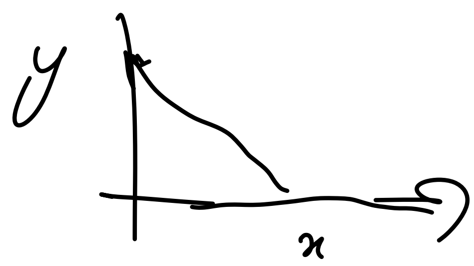
$\boxed{p \geq 1}$

$\boxed{0 < p < 1}$

① Евклидова метрика

$p = 2.$

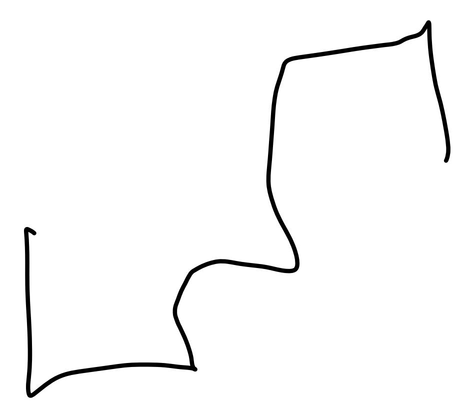
$$\left(\sum |x_j - z_j|^2 \right)^{1/2}$$

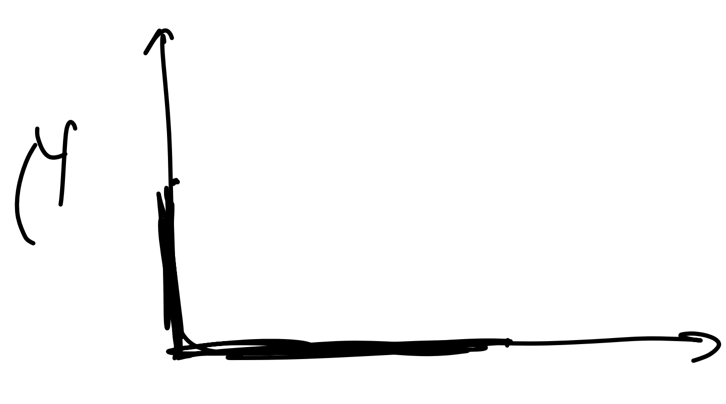


② Манхэт. расст-ие

$p = 1$

$$\sum |x_j - z_j|$$





откуда берется?

③ $p = \infty$

$$\rho_\infty (x, z) = \max_{j=1..d} |x_j - z_j|$$

$p = 0$ — счит. расст-ие

$$\rho_0(x, z) = \sum_{j=1}^{d} [x_j \neq z_j]$$

можно добавлять веса

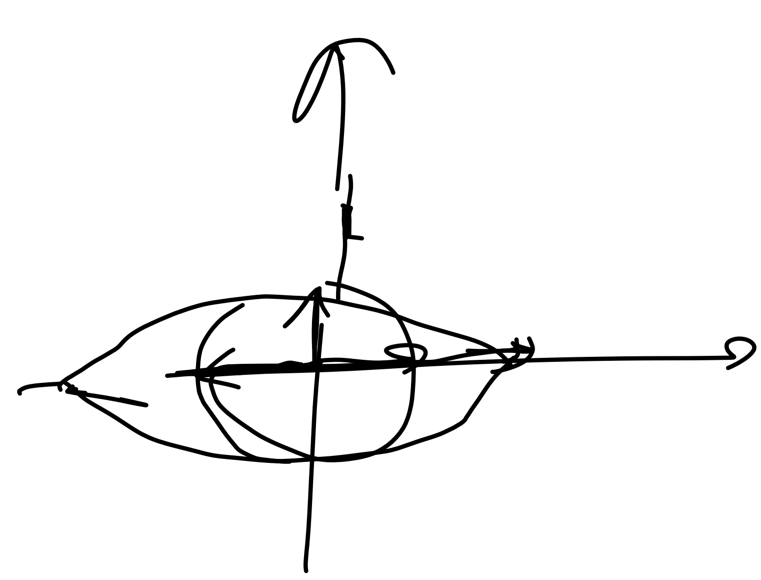
$$\rho_p(x, z; w) = \left(\sum_{j=1}^{d} w_j |x_j - z_j|^p \right)^{1/p}$$

$$f(x) = \rho_2(x, 0; w) = \left(\sum_{j=1}^{d} w_j x_j^2 \right)^{1/2}$$

эллипс

$$f^2(x) = \sum_{j=1}^{d} w_j x_j^2 = \sum \tilde{x}_j^2$$

$$\tilde{x}_j = x_j \cdot \sqrt{w_j}$$



Расст-ие между текстами

① Косинусное расст-ие

	W_1	W_2	W_3	...	W_D
Текст 1	1	0	1		
Текст 2	0	0	1		
Текст 3	0	1	0		
Текст 4	0	1	0		

$$\langle x, z \rangle = \| x \| \| z \| \cdot \cos(\theta)$$

$$\cos(\theta) = \frac{\langle x, z \rangle}{\| x \| \| z \|}$$

$$\cos(\theta_{T1, T3}) = 0.$$

$$\cos(\theta_{T3}, \theta_{T4}) = \frac{1}{1 \cdot 1} = 1$$

$$\cos(\theta_{T1}, \theta_{T2}) = \frac{1}{\sqrt{2 \cdot 1}} = \frac{1}{\sqrt{2}}$$

A, B – мн-ва

$$\rho_J(A, B) = \boxed{1 - \frac{|A \cap B|}{|A \cup B|}}$$

Рассмотрим сами