$< O(l)$

$\ll o(l)$

$\underbrace{\text{LSH}}$, NSW, HNSW

$f : X \to H$ $\qquad$ $J_{A,B}$

$\mathcal{F} : \{ f : X \to H \}$ $\qquad$ $\rho_{x,z}$

Опр. Семейство хэш-функций $\mathcal{F}$ с распределением $P(f)$ называется (d_1, d_2, p_1, p_2) - чувствительным, если:

1) $\rho(x, z) \le d_1 \Rightarrow \mathbb{P}_{f \in \mathcal{F}}[f(x) = f(z)] \ge p_1$

2) $\rho(x, z) \ge d_2 \Rightarrow \mathbb{P}_{f \in \mathcal{F}}[f(x) = f(z)] \le p_2$

$d_1 < d_2 \qquad p_1 > p_2$

Min-hash (Расст. Жаккара)

Объекты - множества

$x \subset U = \{u_1 \dots u_n\}$

π - перестановка на U

π - важность слова по отношению к теме.

π:

0	u_{100}	Гомотопичность
1	u_{321}	Многообразие
2	u_{415}	Интеграл
$\vdots$		
$n-1$	u_{10}	Барби
n	u_{152}	Эскимо

$$f_\pi(A) = \min\{\pi(i) \mid u_i \in A\}$$

f_π(„Борис построил многообр") = 1

f

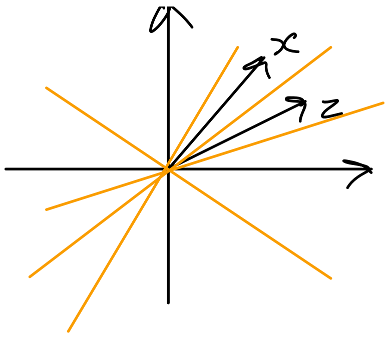
π будет сэмплироваться

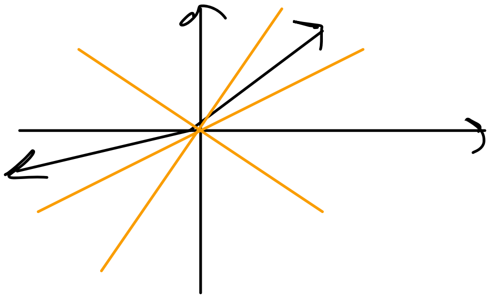
$$\rho(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

Косинусное расстояние

$$\rho(x, y) = \arccos \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

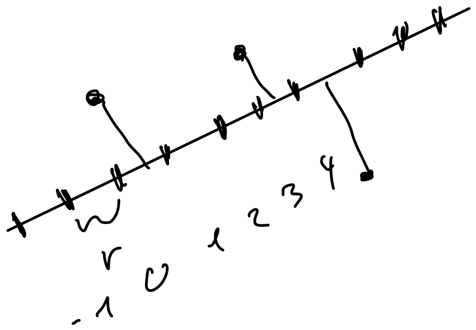
$$\mathcal{F} = \{ f_w(x) = \operatorname{sign} \langle w, x \rangle \mid w \in R^d \}$$





Евклидово расстояние

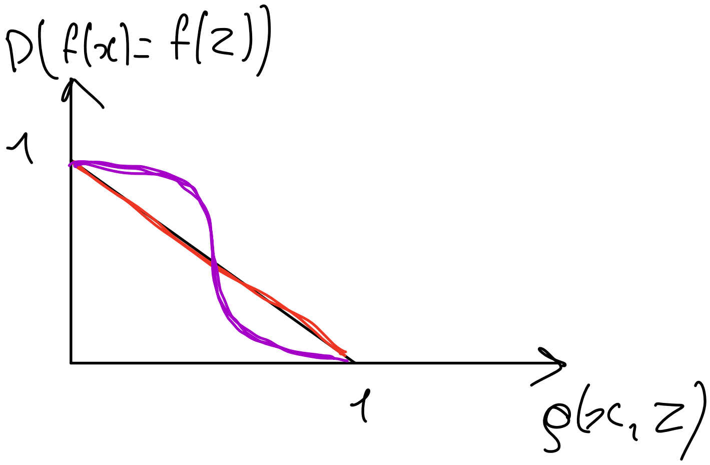
$$\mathcal{F} = \left\{ f_{w,b}(x) = \left\lfloor \frac{\langle w, x\rangle + b}{r} \right\rfloor \;\middle|\; w \in \mathbb{R}^d, b \in [0, r] \right\}$$



$w \sim N(0, I)$

$b \sim U[0, r)$

$\rho \in (0; 2]$



Композиция хэш-функций

$$g(x) = (f_1(x), \dots, f_m(x))$$

$P[f_1(x) = f_1(z)] = p$

$P(g(x) = g(z)) = p^m$

d_1, d_2, p_1, p_2 – чувств

Алгоритм

$$x \downarrow g(x)$$

$N(x) \{x_i \in X \mid g(x_i) = g(x)\}$

$\downarrow$

KNN среди $N(x)$

$$\begin{cases} g_1(x) = [f_{11}(x), \dots, f_{1m}] \\ \vdots \\ g_L(x) = [f_{L1}(x), \dots, f_{Lm}] \end{cases}$$

$$P([g_1(x) = g_1(z)] \,||\, \dots \,||\, [g_L(x) = g_L(z)]) =$$

$$= \underbrace{1-(1-p^m)^L}_{\text{график, который мы хотим}}$$

NSW

Navigable small world

$G = (X, E)$ u – новое наблюдение

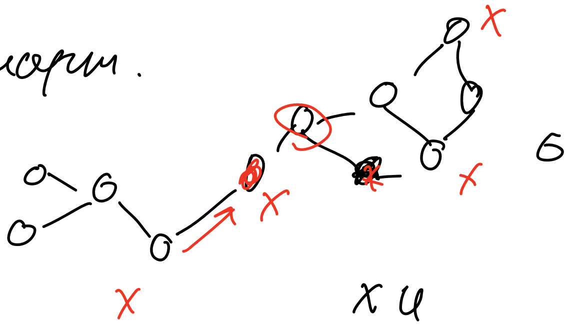
1) Берём случайную вершину v в G

2) В цикле: среди соседей u' ищем:

$$\rho(v', u) < \rho(v, u)$$

Если такой сосед нашёлся, переходим в него

Используем мультистарт.



Добавление вершины u

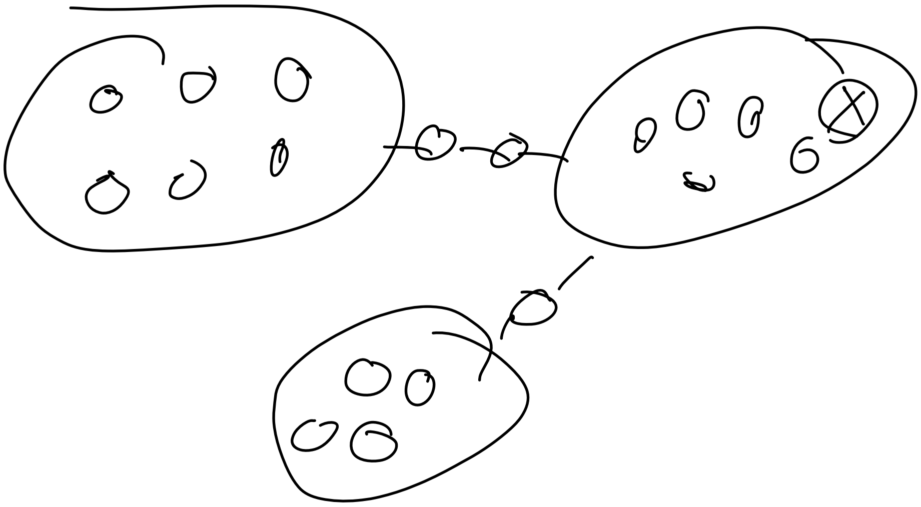
Пусть имеем граф некоторого размера.

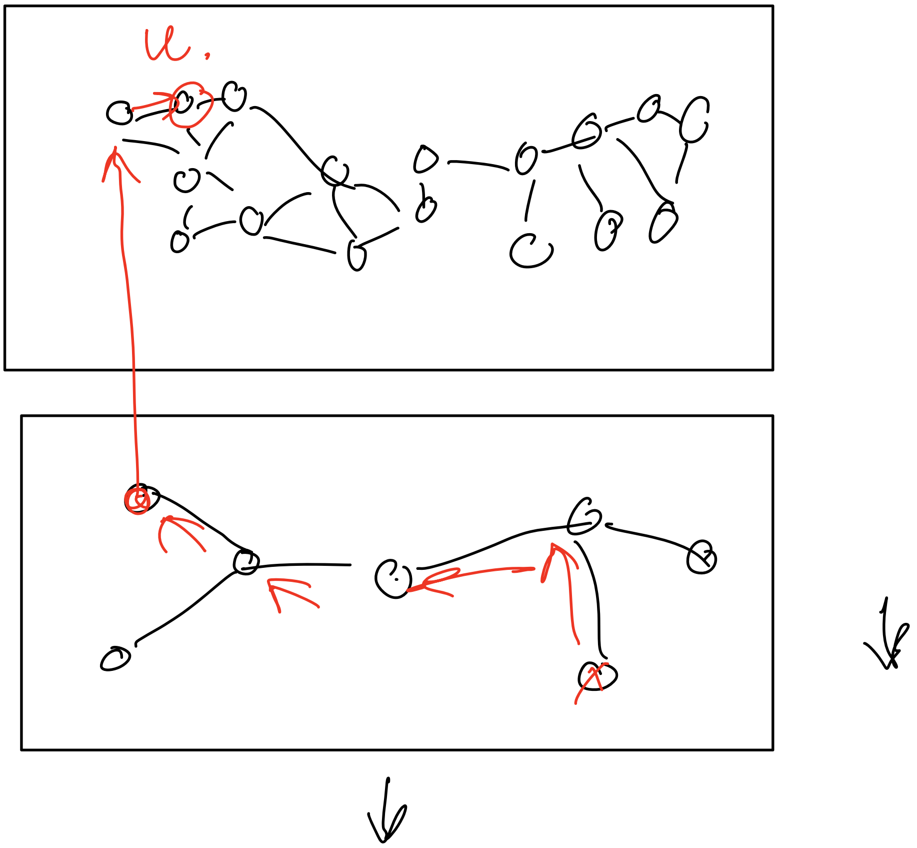
Мультистарт $\rightarrow C(u)$

$D(u) = C(u) \cup \{$ соседние вершины для всех $v \in C(u) \}$

Соединяем u с ближайшими соседями из $D(u)$

HNSW




u.