

# Машинное обучение.

## Домашнее задание №2

**Задача 1.** Покажите, что расстояние Джаккарда является метрикой.

Подсказка: может пригодиться доказанное на занятии утверждение о том, что коэффициент Джаккарда равен кое-какой вероятности.

**Задача 2.** Пусть зависимость вероятности коллизии от расстояния между объектами имеет вид  $p = 1 - d$  (для исходного семейства хэш-функций). После построения композиции над этим семейством вероятность примет вид  $1 - (1 - (1 - d)^m)^L$ . Данное выражение как функция от  $p$  имеет вид сигмоиды. Найдите координату точки ее перегиба, и исследуйте ее поведение при стремлении  $L$  к бесконечности при фиксированном  $m$  (и наоборот, при стремлении  $m$  к бесконечности при фиксированном  $L$ ). Проинтерпретируйте результаты.

**Задача 3.** Рассмотрим универсальное множество слов

$$U = \{\text{зависимость, вероятность, коллизия, испанская, инквизиция, расстояние}\}.$$

Вычислите MinHash для множества

$$A = \{\text{зависимость, вероятность, инквизиция}\}$$

для перестановок  $\pi_1, \pi_2$ :

	зависимость	вероятность	коллизия	испанская	инквизиция	расстояние
$\pi_1$	4	2	6	1	3	5
$\pi_2$	3	6	2	5	6	1