

Исследование поведения характеристик графов при различии распределений

29 мая 2025 г.

Часть 1.

Отчет Самсанович Екатерины

- Гипотеза $H_0: X_i \sim N(0, \sigma_0)$
- Гипотеза $H_1: X_i \sim t(\nu_1)$

Типы графов:

- KNN-граф: максимальная степень
- DIST-граф: размер максимального независимого множества

1. Фиксированная выборка, изменяемые параметры распределений

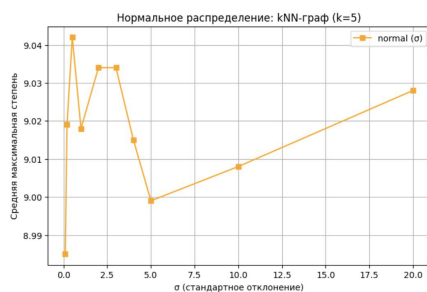


Рис. 1: Normal(0,σ), knn

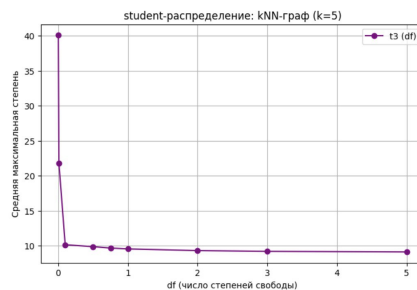


Рис. 2: ST, knn

- Значение $max - degree$ практически не меняется в зависимости от σ и держится в узком диапазоне около 9. Можно сделать вывод, что варьирование стандартного отклонения в рассматриваемом диапазоне не сильно меняет плотность соседей в kNN-графе.

Стандартное нормальное распределение с различными σ создает более однородный набор данных без резких выбросов, что отражается в стабильности $max - degree$.

- При очень маленьком значении $\alpha = 0.001$ наблюдается очень высокая средняя максимальная степень (около 40). По мере увеличения α значение $max - degree$ резко падает и стабилизируется примерно в диапазоне 9–10 при α от 0.5 и выше. Эта динамика отражает влияние степени свободы t-распределения на разброс данных: при увеличении α распределение приближается к нормальному, что приводит к уменьшению разброса и стабилизации максимальной степени.

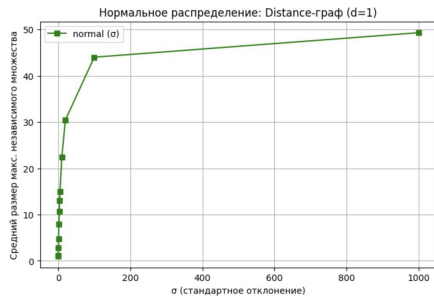


Рис. 3: Normal(0,σ), dist

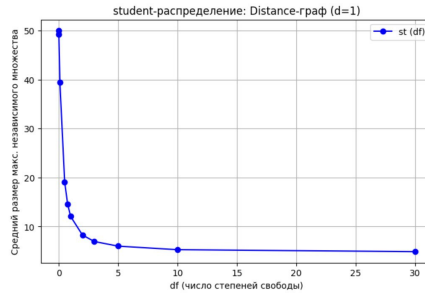


Рис. 4: ST, dist

- При малых значениях σ размер MIS очень маленький (порядка 1-2), что говорит о том, что большинство точек близко друг к другу. По мере увеличения σ размер MIS растёт практически линейно и достигает значения близкого к размеру выборки (около 49 при $\sigma = 1000$). Это говорит о том, что точки с большим σ разбросаны сильнее, DIST-граф становится более разреженным, и максимальное независимое множество увеличивается.
- При очень маленьких значениях α (0.001, 0.01) размер MIS практически равен размеру выборки (около 50). По мере увеличения α размер MIS резко уменьшается и стабилизируется на низких значениях (около 5 при $\alpha = 10$ и выше). Это означает, что при малых степенях свободы student-распределения точки разбросаны сильно — между ними мало ребер в DIST-графе (из-за фиксированного порога расстояния), что делает множество независимым большим. При увеличении α распределение становится ближе к нормальному, точки сгущаются, увеличивается число ребер, как итог — максимальное независимое множество становится значительно меньше.

2. Фиксированные параметры распределений, изменяемая выборка

- При увеличении максимальной степени распределения ведут себя почти одинаково, немного выше распределение student.
- Что касается MIS, там все более однородно, а при $k=500$ вообще постоянно.

3.

- На KNN получились неразделимые данные.
- Здесь достаточно визуально разделимыми данные становится при $n = 1200$, при этом можно увидеть, что на многих графиках мощность равна 1, ошибка не превысила допустимого значения 0.055