# Задача 12

Предсказать сорт винограда, из которого сделано вино, используя результаты химических анализов , с помощью KNN — метода k ближайших соседей с тремя различными метриками. Построить график зависимости величины ошибки от числа соседей $k$.

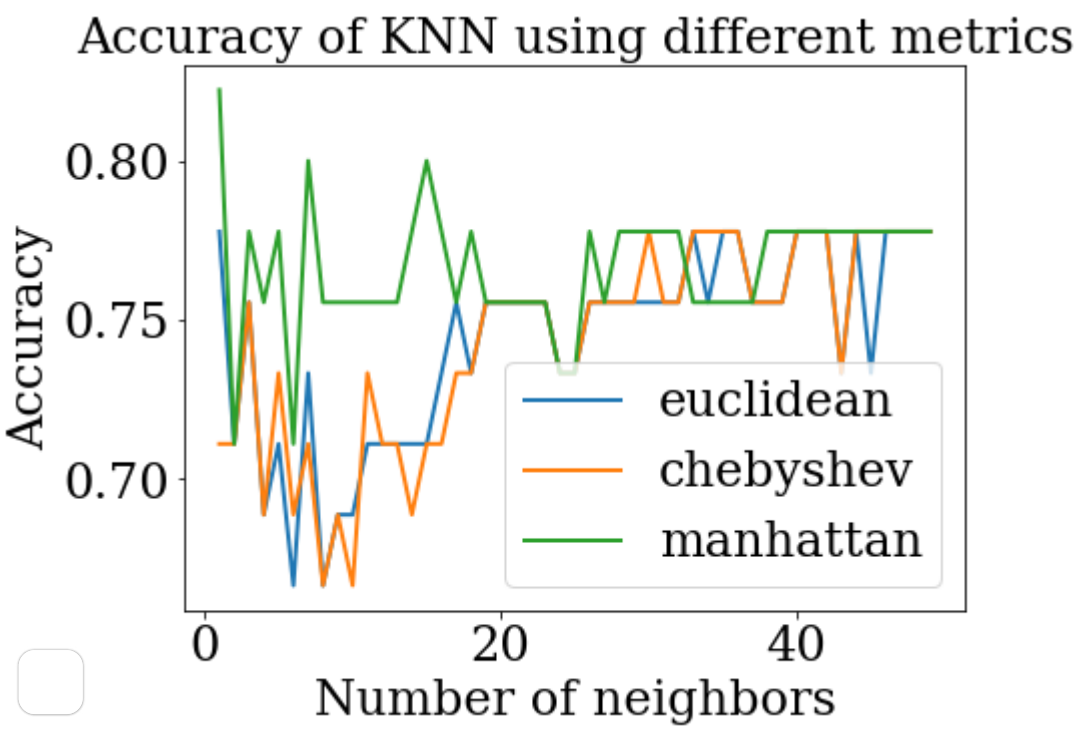
Загрузим датасет с информацией о винах:

Out[2]:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 14.23 | 1.71 | 2.43 | 15.6 | 127 | 2.80 | 3.06 | 0.28 | 2.29 | 5.64 | 1.04 | 3.92 | 1065 |
| 1 | 1 | 13.20 | 1.78 | 2.14 | 11.2 | 100 | 2.65 | 2.76 | 0.26 | 1.28 | 4.38 | 1.05 | 3.40 | 1050 |
| 2 | 1 | 13.16 | 2.36 | 2.67 | 18.6 | 101 | 2.80 | 3.24 | 0.30 | 2.81 | 5.68 | 1.03 | 3.17 | 1185 |
| 3 | 1 | 14.37 | 1.95 | 2.50 | 16.8 | 113 | 3.85 | 3.49 | 0.24 | 2.18 | 7.80 | 0.86 | 3.45 | 1480 |
| 4 | 1 | 13.24 | 2.59 | 2.87 | 21.0 | 118 | 2.80 | 2.69 | 0.39 | 1.82 | 4.32 | 1.04 | 2.93 | 735 |

Загруженные данные не содержат пропусков, признакии являются действительными числами, поэтому их можно использовать для обучения модели без предварительной обработки.

Реализуем KNN. Из опсания данных видно, что классы достаточно сбалансированы, поэтому в качестве метрики качества будем использовать accuracy. В качестве метрик для классификатора будем использовать евклидову, манхэттенскую и чебышевсткую метрики, предварительно зафиксировав `random_state` для воспроизводимости результатов:




Из графика видно, что евклидова и чебышевская метрика в зависимости от числа соседей ведут себя

схожим образом, в то время как манхэттенская метрика почти при всех значениях $k$ показывает несколько лучший результат. Другое наблюдение, касающееся манхэттенской метрики, заключается в том, что значение accuracy при ее использовании слабо зависит от $k$, в отличие от евклидовой и чебышевской метрик, имеющих явный спад на графике при $k$ примерно равном $10$.

Таким образом, из графика можно сделать вывод, что хотя все три метрики показывают в целом схожие значения точности, наилучший результат `accuracy = 0.82` можно получить, используя манхэттенскую метрику при $k = 1$ (т. е. метод ближайшего соседа).