Необходимо разделить точки разных классов на плоскости таким образом, чтобы максимизировать расстояние от них до делящей линии (в общем случае в многомерном пространстве строится гиперплоскость).

Используется библиотека sklearn.svm.SVC, подбираются параметры функции ядра (['linear', 'poly', 'rbf', 'sigmoid']) и силы регуляризации (а также некоторые другие специфичные параметры для каждого ядра). Как обычно, осуществляется перебор доступных вариантов, результаты сравниваются (для каждого результата считается средняя f1\_score по всем cross-validation - разбиениям, выбирается результат с большей средней ф-мерой)

Полученные результаты-

{

'chips': {

'regularization\_strength': 10.0,

'kernel': 'rbf',

'gamma': 'scale',

'filename': 'data/chips.csv'

},

'geyser': {

'regularization\_strength': 10.0,

'kernel': 'poly',

'gamma': 'scale',

'degree': 3,

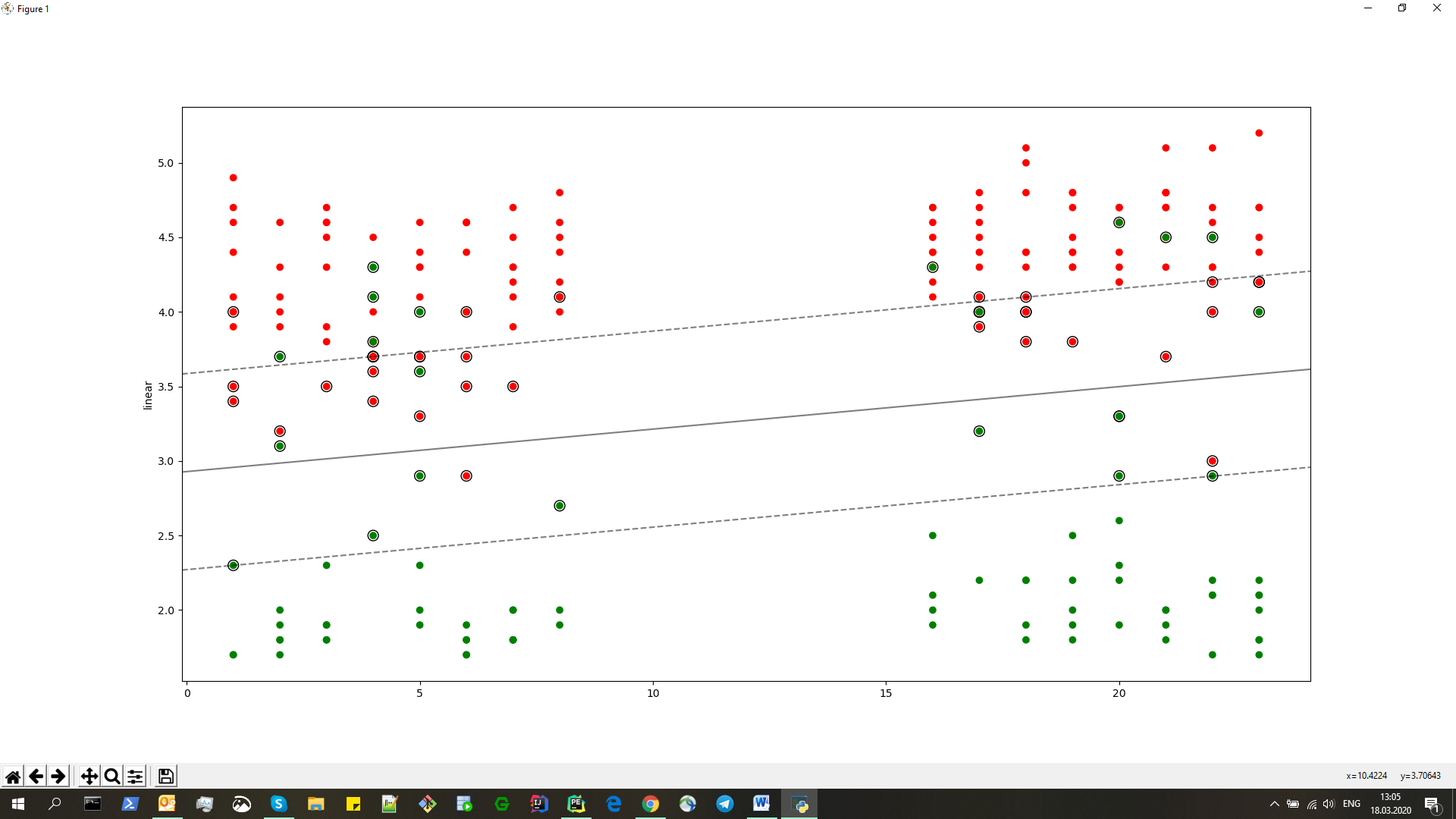
'filename': 'data/geyser.csv'

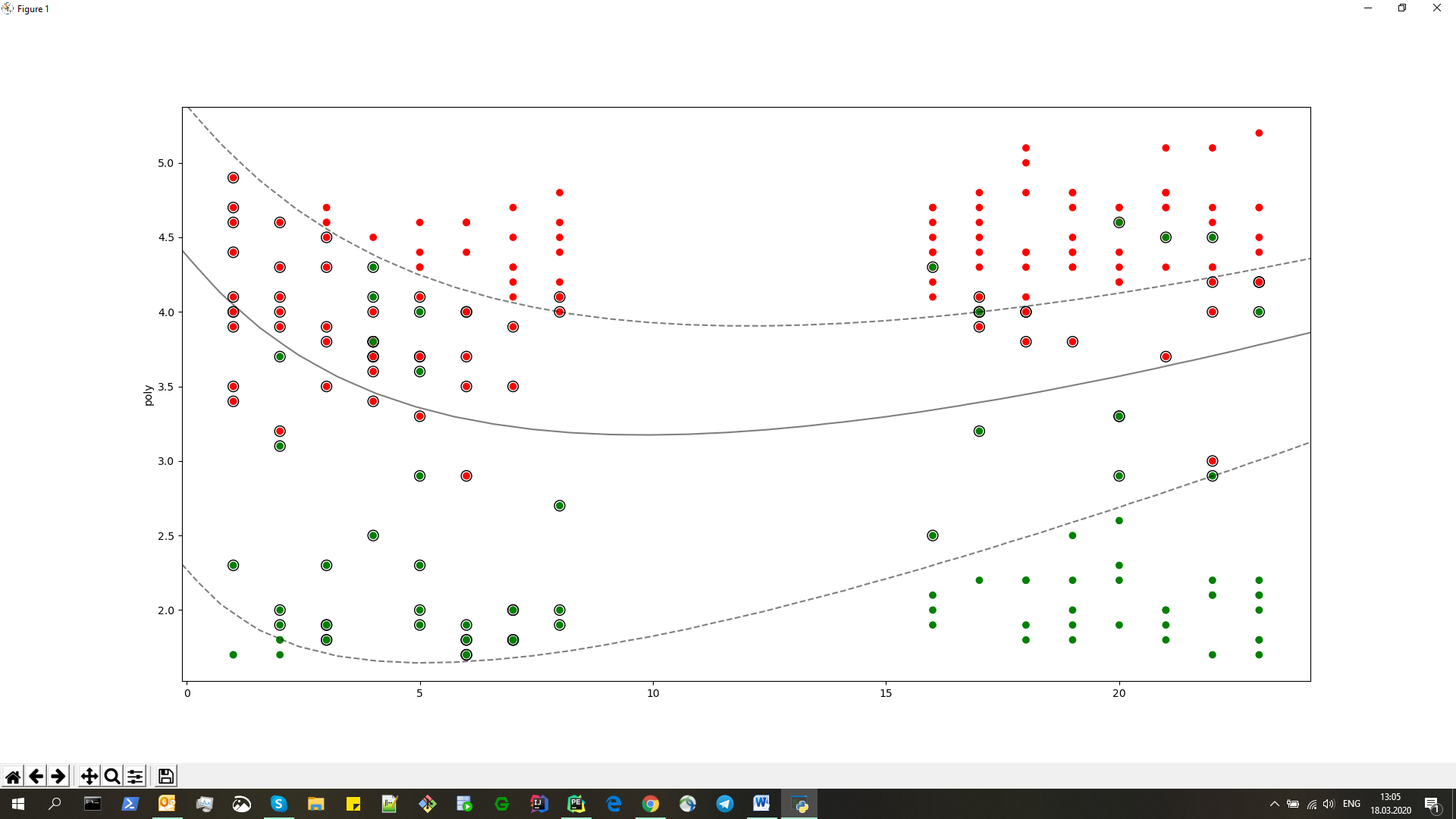
}

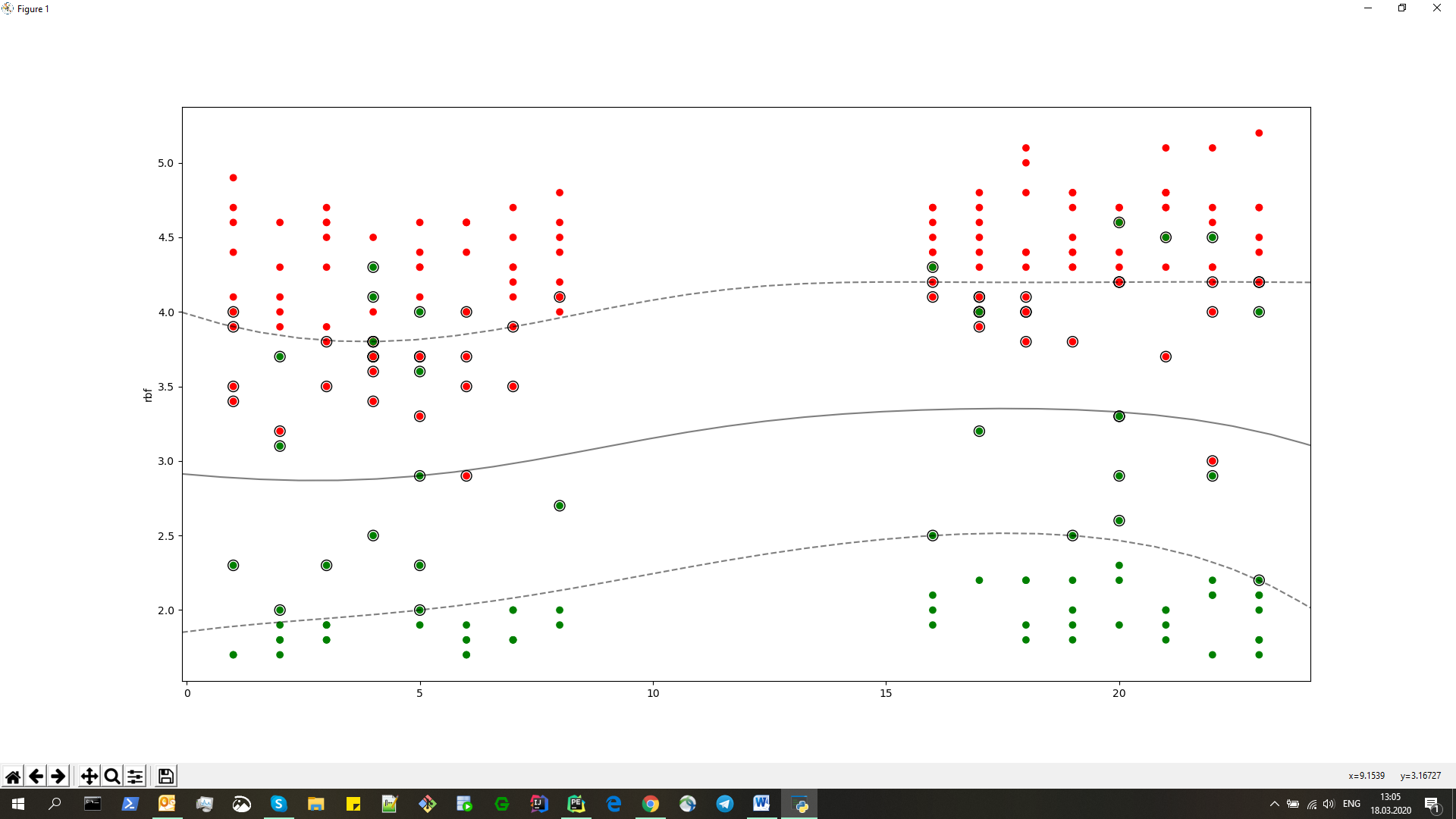
}

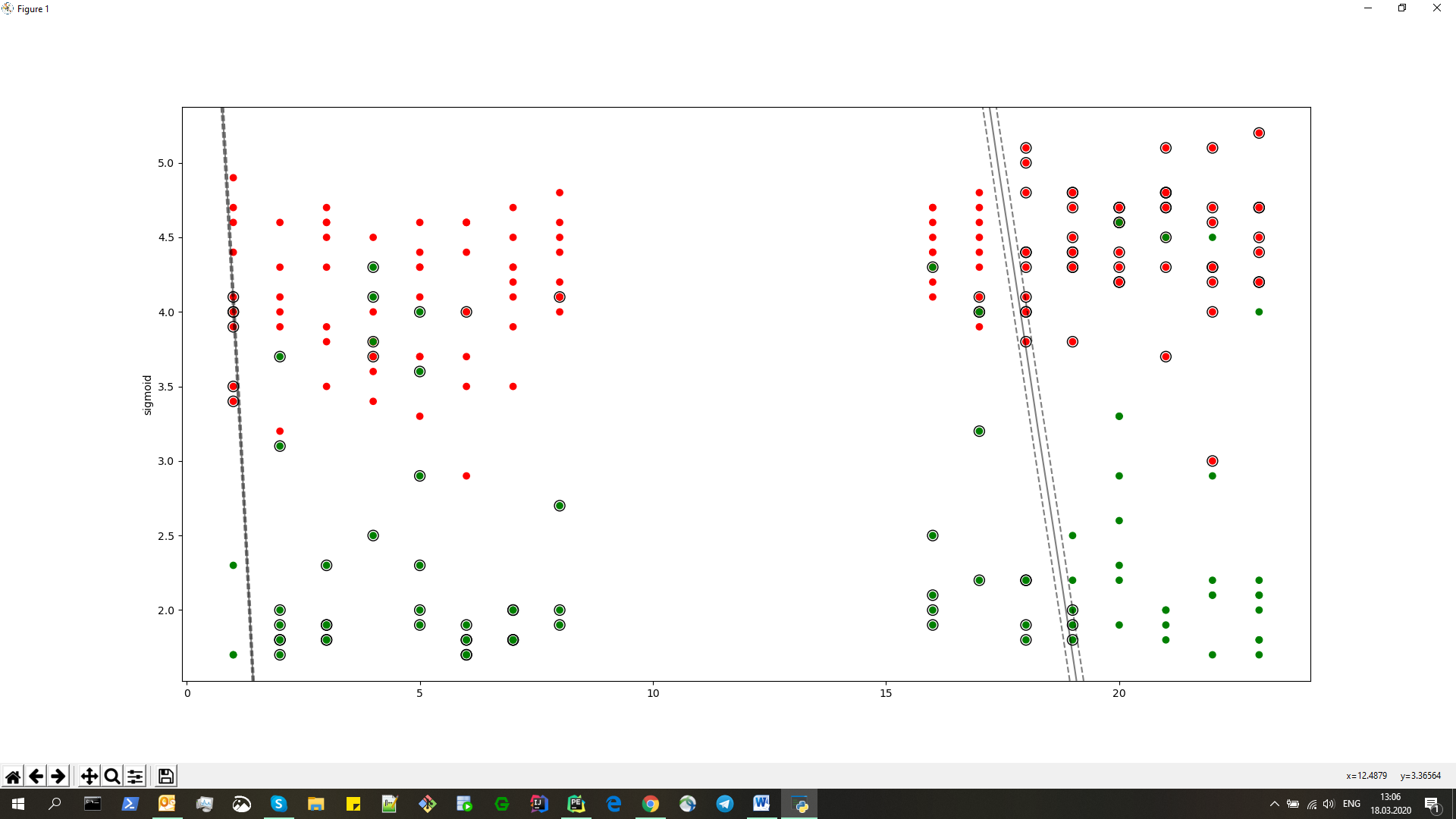
Графики:

'geyser'









'chips'

