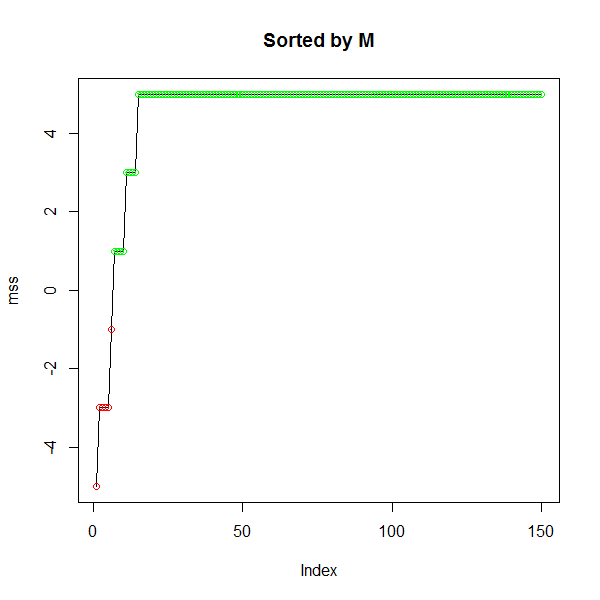
Тема: Алгоритм STOLP. Балицкая Анастасия 401-И

В выборке есть выбросы и эталонные элементы. Алгоритм STOLP позволяет выбросить ошибочные элементы и взять эталонные. Определяет он «эталонность» объекта по функции M. Посмотрим на отсортированный список значений M(x) для всех элементов:



Элементы помеченные красным – выбросы. Алгоритм удалит их на самой начальной стадии:

good **=** rep**(**T, times**=**l**)**

**for** **(**i **in** 1**:**l**)** **{**

print**(**sprintf**(**"removing bad: %d", i**))**

**if** **(**M**(**xl**[**i,1**:**n**]**, xl**[**i,n**+**1**]**, xl**[-**i,**]**, k**)** **<** delta**)** **{**

good**[**i**]** **=** F

**}**

**}**

xl **=** xl**[**good,**]**

l **=** dim**(**xl**)[**1**]**

После чего в выборку войдут самые эталонные объекты с максимальным отступом:

inomega **=** rep**(**F, times**=**l**)**

classes **=** length**(**levels**(**xl**[**,n**+**1**]))**

idxval **=** matrix**(**0, nrow**=**classes, ncol**=**2**)**

**for** **(**i **in** 1**:**l**)** **{**

print**(**sprintf**(**"inital omega: %d", i**))**

curr **=** M**(**xl**[**i,1**:**n**]**, xl**[**i,n**+**1**]**, xl**[-**i,**]**, k**)**

cls **=** xl**[**i,n**+**1**]**

**if** **(**idxval**[**cls,2**]** **<** curr**)** **{**

idxval**[**cls,1**]** **=** i

idxval**[**cls,2**]** **=** curr

**}**

**}**

И потом присоединяются объекты, на которых происходит ошибка:

omega **=** xl**[**inomega,**]**

xl\_omega **=** xl**[!**inomega,**]**

worst **=** matrix**(**1e9, nrow**=**l, ncol**=**2**)**

**for** **(**i **in** 1**:**l**)** **{**

print**(**sprintf**(**"omega length: %d, current i: %d", length**(**inomega**[**inomega**==**T**])**, i**))**

**if** **(!**inomega**[**i**])** **{**

val **=** M**(**xl**[**i,1**:**n**]**, xl**[**i,n**+**1**]**, omega, k**)**

**if** **(**val **<** delta**)** **{**

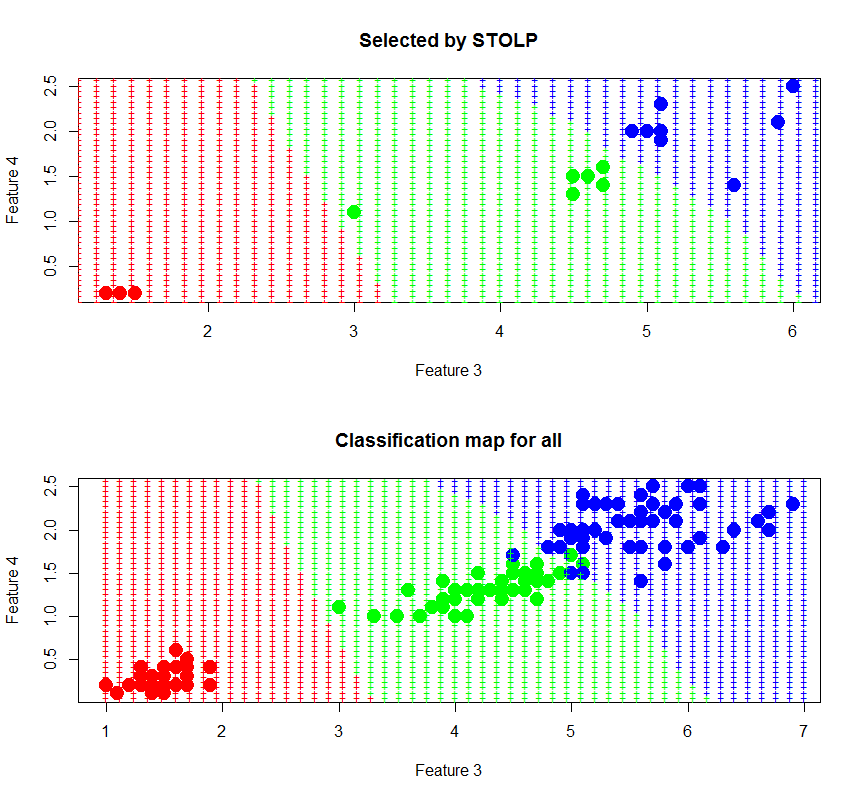
worst**[**i,**]** **=** c**(**i, val**)**

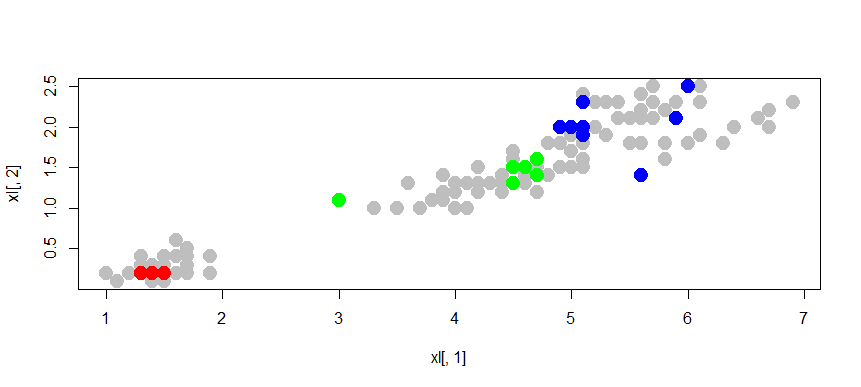
**}**

**}**

**}**

Посмотрим на выбранные объекты:





Видно что алгоритм правильно выбрал эталоны и граничные объекты.

Сравним точность на полной выборке и выборке STOLP:

Полная выборка: 4%

Выборка STOLP: 4%

Погрешность не уменьшилась, но поскольку алгоритм оставил только 19 объектов то kNN будет работать в 8 раз быстрее.

Вывод: В выборке есть эталоны и выбросы, чтобы удалить выбросы и взять эталоны и граничные объекты мы используем алгоритм STOLP. Поскольку остается меньше объектов метрические алгоритмы классификации работают быстрее и есть шанс, что ошибка уменьшится.