Сафиуллин Амир

Решение БДЗ содержит следующие методы:

Алгоритм 1. vlob(plus, minus, x\_test, y\_test)

Алгоритм основан на нормированной сумме мощности пересечения признаков неизвестного примера с примерами-(+) и примерами-(-).

То же самое для отрицательных.

Неизвестный пример относится к тому набору, где эта сумма больше.

Алгоритм 2. is\_in\_intent(plus, minus, x\_test, y\_test)

Пересекаем с положительным и проверяем чтобы пересечение не вкладывалось ни в одно отрицательное. если все так, то начисляем голос в виде "относительной мощности пересечения".

То же самое для отрицательных.

Где сумма накопленных "голосов" больше - туда и классифицируем, в случае равенства смотрим по поддержке, как в алгоритме 3.2 (с порогом 10.)

Алгоритм 3.1. is\_in\_int1(plus, minus, x\_test, y\_test)

Для неизвестного примера начисляем голос пропорционально "относительной мощности пересечения" пересечения (признаков неизвестного примера и (+)примера) с остальными (+)-примерами, если это пересечение вкладывается в них.

Для минуса то же самое.

Пример классифицируется туда, где сумма «голосов» больше.

Алгоритм 3.2. is\_in\_int2(plus, minus, x\_test, y\_test)

Лучший порог = 10; лучший порог искался с помощью скользящего контроля максимизуруя accuracy

Для неизвестного примера начисляем голоса пропорционально поддержке в положительных примерах и мощности пересечения с положительным примером, если она больше порога.

Для минуса то же самое.

Пример классифицируется туда, где сумма «голосов» больше.

Алгоритм 4. costs\_and\_penalty(plus, minus, x\_test, y\_test, tresh)

Объединение алгоритмов 2 и 3.1. Начисляем голоса, вдобавок задаем штраф (отнимаем голоса) для неизвестного примера если пересечение с плюс примером вкладывается в отрицательный пример и штраф если пересечение с минус примером вкладывается в положительный.

Классифицируем туда, где сумма голосов больше.

Ниже приведена таблица со средними значениями accuracy для каждого алгоритма (1-4), и для трех наиболее популярных — SVC, Randome forests, k-Nearest Neighbor.

|  |  |
| --- | --- |
| **Алгоритмы** | **Среднее значение по accuracy (на tic tac)** |
| 1 | 0,659 |
| 2 | 0,99 |
| 3.1 | 0,933 |
| 3.2 | 0,979 |
| 4 | 0,989 |
|  |  |
| SVC | 0,998 |
| Randome Forests | 0,986 |
| kNN | 0,982 |

На дата сете Kr vs kp:

Accuracy алгоритма 1 — 0,799

Accuracy алгоритма 2 — 0,979