

## Описание алгоритмов.

### Alg1.

Он классифицирует простым относительным большинством — объект, который нужно классифицировать, пересеем с объектом из +контекста. Потом бежим по всем объектам из -контекста, и смотрим, вкладывается ли описание пересечения в описание объекта из -контекста. Если вкладывается, то ничего не делаем, а если не вкладывается ни в один объект из -контекста, то рассматриваемый объект из +контекста голосует за принадлежность классифицируемого объекта к положительному контексту. То же самое проедем с объектами из -контекста. Потом считаем относительное число голосов за + и за -, и сравниваем. Если за + больше, то пусть это будет положительный объект, и наоборот. Точность тут очень хорошая — на тестовых примерах она была около 98%. Но работает программа ужасно медленно.

### Alg2.

Он смотрит на размер пересечения классифицируемого объекта и объектов из + и — контекстов. Если это пересечение больше 5, то объект считается «похожим» на объект из + или - контекста, и тот объект из + (или - контекста), с которым пересекались, голосует за принадлежность к +(или -). Несмотря на свою наивность, на тестовых примерах работал достойно, с точностью около 80%.

### Alg3.

Это линейная комбинация первого и второго алгоритмов. Мы смотрим на пересечение с объектом из +контекста, и если это пересечение больше 4, то прогоняем по -контексту в поисках вложения и уже тогда голосуем. А если меньше, то ничего не делаем.

Так мы получаем гораздо быстрее работающую модификацию алгоритма 1, хотя и теряем в точности — она теперь около 95%.

### Alg4.

Все предыдущие алгоритмы были достаточно незамысловаты. Ведь может же быть признак, который достаточен для вхождения в тот или иной контекст, а значения всех остальных признаков вообще не важны (я говорю, что такое может быть, а не что это обязательно есть в наших данных). И алгоритмы 3 и 2 в такие вещи не учитывают. Поэтому в алгоритме 4 происходит следующее:

1. +контекст делится на группы по 10 примеров.
2. В этих группах ищутся гипотезы
3. Все эти гипотезы проходят проверку на фальсификацию
4. Все то же самое продлевается для -контекста

У нас теперь есть то, что можно назвать предположительно важными группами значений признаков, которые отвечают за принадлежность к тому или иному классу. Делила на группы по 10 для того, чтобы мы закончили считать до конца следующего года.

5. Идем по списку гипотез, и смотрим, вкладываются ли они в описание классифицируемого объекта. И берем такую же схему голосования, как в первом, только теперь голосуют гипотезы.