

ЗАДАНИЕ ДЛЯ 317 ГРУППЫ

Написать алгоритм, который находит оптимальное признаковое пространство в задаче классификации для метода ближайшего соседа с косинусной мерой сходства. Функционал качества метода – процент правильных ответов на скользящем контроле (leave-one-out) – его и надо максимизировать.

Замечание. Если в признаковом пространстве косинусная мера определена некорректно (это происходит, когда какая-то подстрока становится нулевой), то функционал качества не определен.

Параметры задач

Все задачи – задачи классификации с двумя классами («ноль» и «один»).

	Число строк	Число столбцов	Ориентировочное качество	Процент первого класса	Проверка 1:10
task1.txt	10	30	0.7	40	0.5
task2.txt	100	50	0.78	50	0.63
task3.txt	100	300	1	54	0.52

Формат исходных файлов:

```
1,13.07,9.9,10.01,10.01
0,8.08,12.07,19.99,19.99
1,15.11,17.88,99.4,99.4
```

Такой файл был бы для задачи с числом строк 3, числом столбцов 4. Первый столбец – целевой (метка классов «0 или 1»).

В качестве ответа студент к 01.10.2012 присылает **на секретный ящик** три текстовых файла

Пример файла для признакового пространства {5, 7, 13, 18}:

```
5 7 13 18
```

Имена файлов должны быть вида **фамилия_№задачи.txt**. Примеры:

Ivanov_1.txt

Ivanov_2.txt

Ivanov_3.txt

Также в теле письма надо указать полученные значения функционала качества для всех задач. Пример:

1 – 0.6

2 – 0.71

3 – 0.99

01.10.2012 состоятся публичные выступления студентов с отчётами по проделанной работе.