

Машинное обучение и интеллектуальный анализ данных

Семинар 5 Лабораторная работа 3

**Г.А. Ососков*, О.И. Стрельцова*, Д.И. Пряхина*,
Д.В. Подгайный*, А.В. Стадник*, Ю.А. Бутенко***

Государственный университет «Дубна»

***Лаборатория информационных технологий, ОИЯИ
Дубна, Россия**

Государственный университет «Дубна»

(Из [1]) Специальное исследование показало, что склонность фирм к утаиванию части своих доходов (и, соответственно, - к уклонению от уплаты части налогов) в существенной мере определяется двумя показателями:

- x_1 – соотношением «быстроты активов» и текущих пассивов;
- x_2 – соотношением прибыли и процентных ставок

(оба показателя оцениваются по определенной методике в шкале от 300 до 900 баллов);

Требуется

- Определить (используя 3 ранее изученных метода классификации), к какому классу (1 или 2) следует отнести фирму с показателями:

$$x_1 = 740, x_2 = 590$$

Провести исследование результатов работы алгоритмов классификации в зависимости от параметров используемых методов.

Сравнить результаты полученные тремя методами, сделать вывод.

Предложить (с обоснованием) наиболее подходящий метод решения задачи.

- Предложить (с обоснованием) и вычислить расстояние, измеряющее степень различия двух анализируемых классов

Алгоритмы классификации:

- Метод опорных векторов
- Наивный Байесовский классификатор
- Метод К-ближайших соседей

№	Фирмы, уклоняющиеся от налогов		Фирмы, не уклоняющиеся от налогов	
	x_1	x_2	x_1	x_2
1	740	680	750	590
2	670	600	360	600
3	560	550	720	750
4	540	520	540	710
5	590	540	570	700
6	590	700	520	670
7	470	600	590	790
8	560	540	670	700
9	540	630	620	730
10	500	600	690	840
11	-	-	610	680
12	-	-	550	730
13	-	-	590	750

Обучающая выборка [1]

Содержание отчета

1. Постановка задачи
2. CSV-файл с исходными данными, представленный в виде таблицы (библиотека *pandas*), рисунок с представленными данными (см. пример)
3. Описание трех методов решения задачи классификации.
4. Исследование результатов работы алгоритмов классификации в зависимости от параметров используемых методов.
5. Метод вычисления расстояния, измеряющий степень различия двух анализируемых классов.
6. Результаты и основные выводы.
7. Список литературы.

