МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Информационные технологии и прикладная математика»

Кафедра 810 «Информационные технологии в моделировании и управлении»

**Лабораторная работа №2**

**по курсу «Интеллектуальный анализ данных»**

**Логистическая регрессия. Метод опорных векторов. Калибровочные кривые. Отбор признаков.**

Выполнил: А.С.Бобряков

Группа: М8O-103М-19

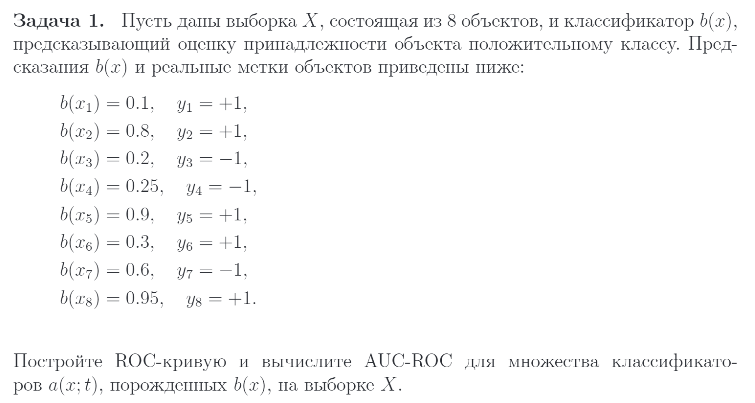
Преподаватель: К.К.Абгарян

Москва, 2020

**Задание**

Обучить метод SVM. Произвести калибровку вероятностей. Сравнить вероятности, полученные методом SVM и логистической регрессией;

**Ход выполнения работы**

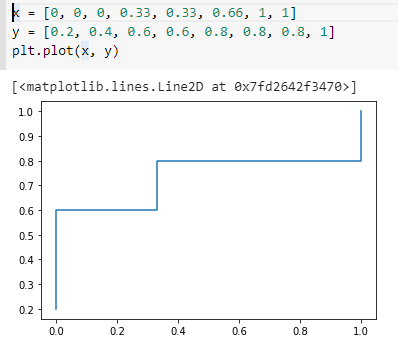


**Решение:**

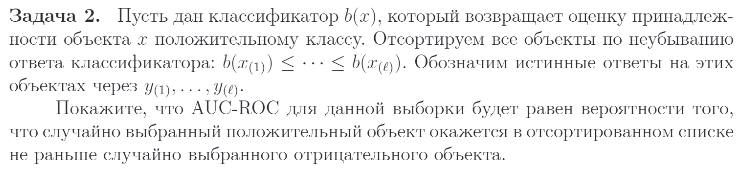
Упорядочим объекты:

|  |  |  |
| --- | --- | --- |
| 1 | 0.95 | 1 |
| 2 | 0.9 | 1 |
| 3 | 0.8 | 1 |
| 4 | 0.6 | -1 |
| 5 | 0.3 | 1 |
| 6 | 0.25 | -1 |
| 7 | 0.2 | -1 |
| 8 | 0.1 | 1 |

ROC – кривая:

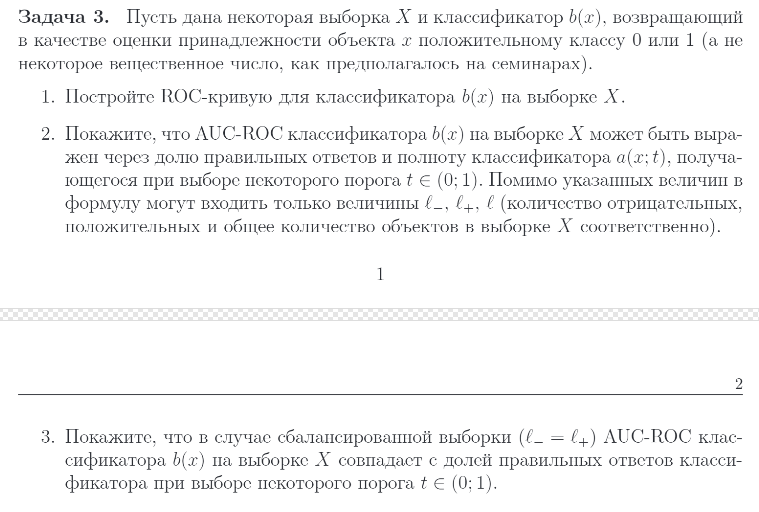


AUC - ROC = 0.732



**Решение:**

Самому вывести тождество не вышло, но нашел статью, где его выводят: <https://pubs.rsna.org/doi/pdf/10.1148/radiology.143.1.7063747>

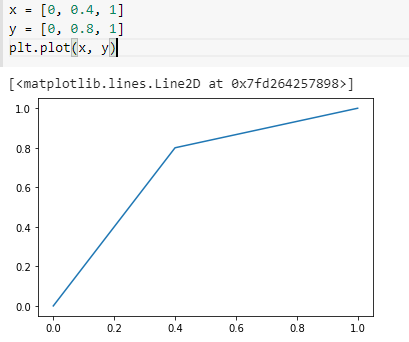


**Решение:**

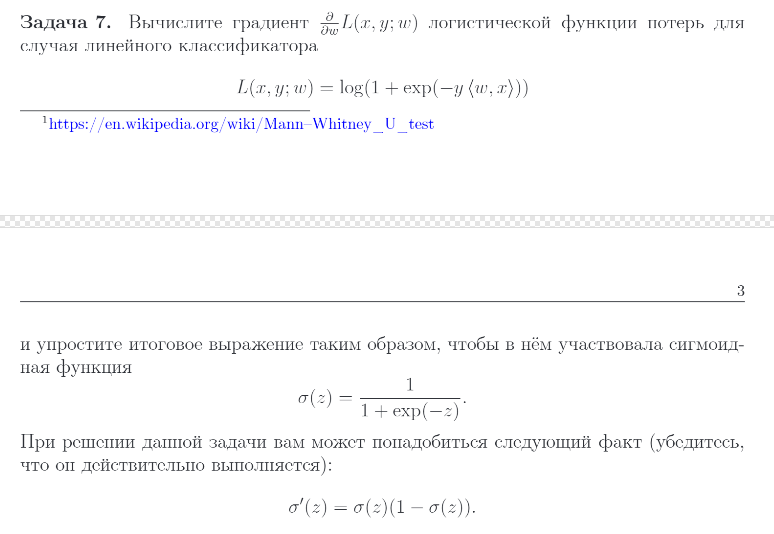
Возьмем выборку из задания 1 с порогом 0,8:

|  |  |  |  |
| --- | --- | --- | --- |
| № | Вероятность | Бинарное значение с порогом 0,4 | Верный ответ |
| 1 | 0.95 | 1 | 1 |
| 2 | 0.9 | 1 | 1 |
| 3 | 0.8 | 1 | 1 |
| 4 | 0.6 | 1 | -1 |
| 5 | 0.3 | -1 | 1 |
| 6 | 0.25 | -1 | -1 |
| 7 | 0.2 | -1 | -1 |
| 8 | 0.1 | -1 | 1 |

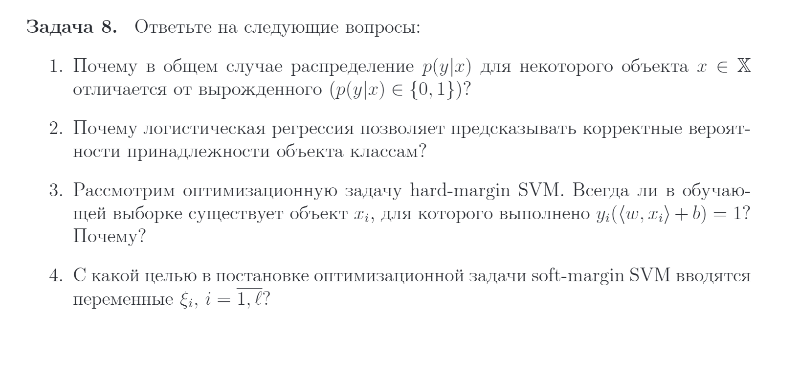
1. ROC – кривая:



2 и 3) Из графика ROC – кривой зависимости TPR от FPR:

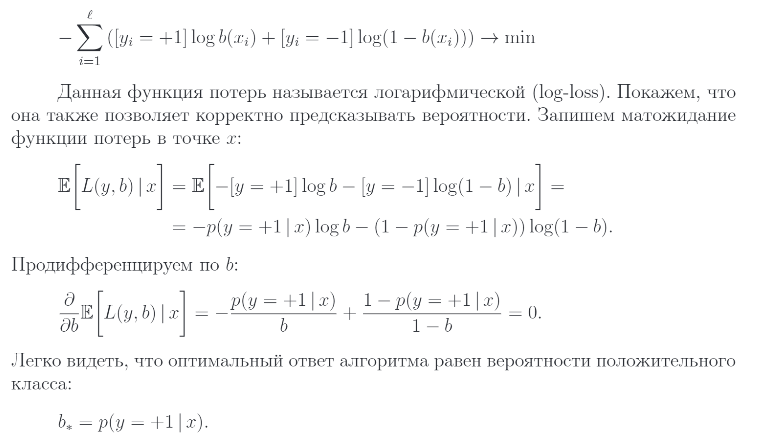


**Решение:**



**Решение:**

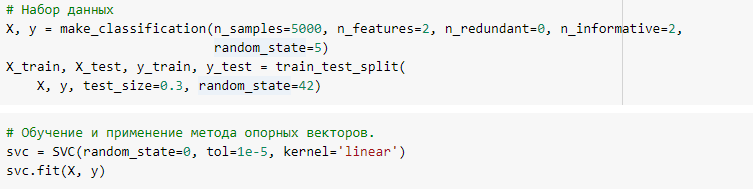
1. -



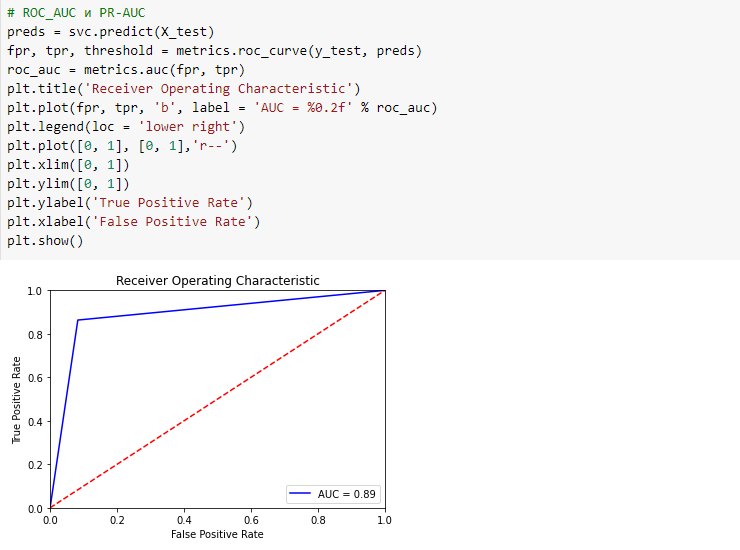
1. Да, потому что это условие соответствует принадлежности точки к полосе разделения. А сам алгоритм SVM максимизирует зазор между гиперплоскостью и объектами классов, которые расположены ближе всего к ней. При этом в каждом из классов найдётся хотя бы один "граничный" объект обучающей выборки, отступ которого равен этому минимуму: иначе можно было бы сместить гиперплоскость в сторону класса с большим отступом, тем самым увеличив минимальное расстояние от гиперплоскости до объектов обучающей выборки.
2. Эти переменные характеризуют величину ошибки на каждом объекте x, чтобы позволить алгоритму допускать ошибки на обучающих объектах, чтобы решить задачу с линейно неразделимыми данными.

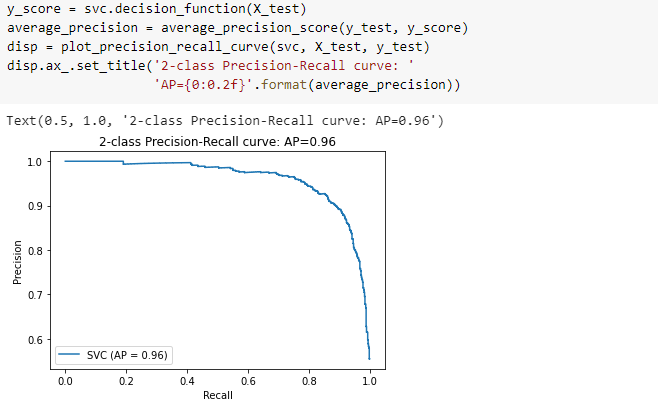
Лабораторная работа.

**Задание**: Обучить метод SVM. Произвести калибровку вероятностей. Сравнить вероятнсти, полученные методом SVM и логистической регрессией

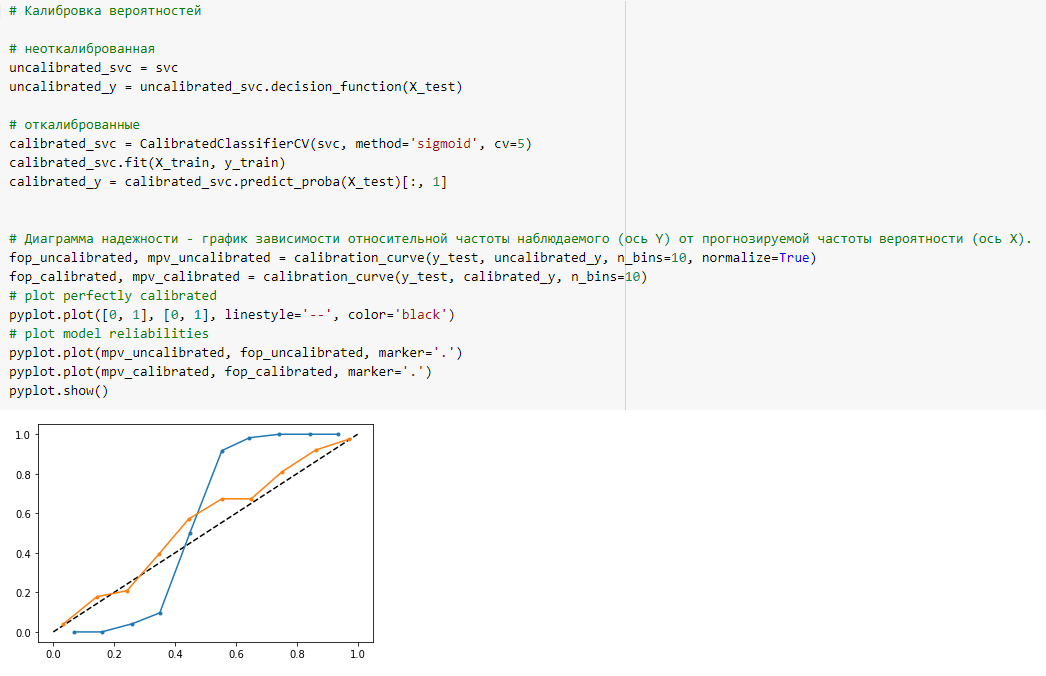


Графики ROC, PR:

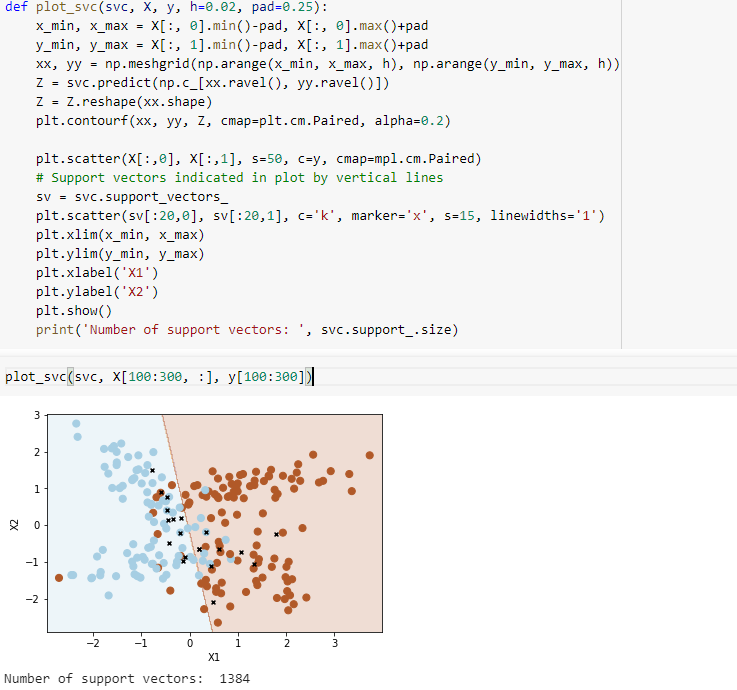




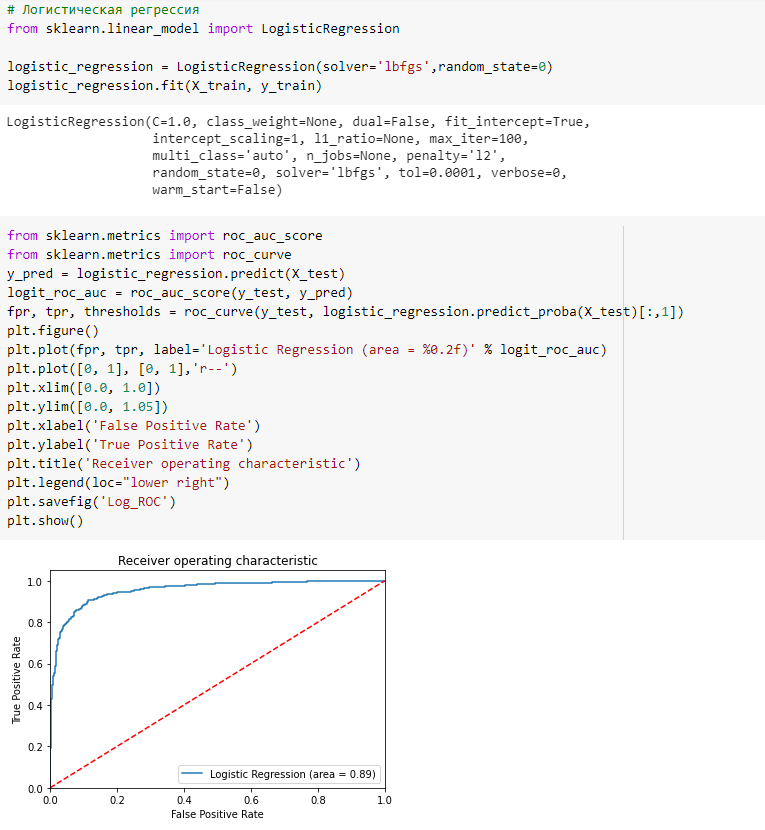
Калибровка вероятностей:



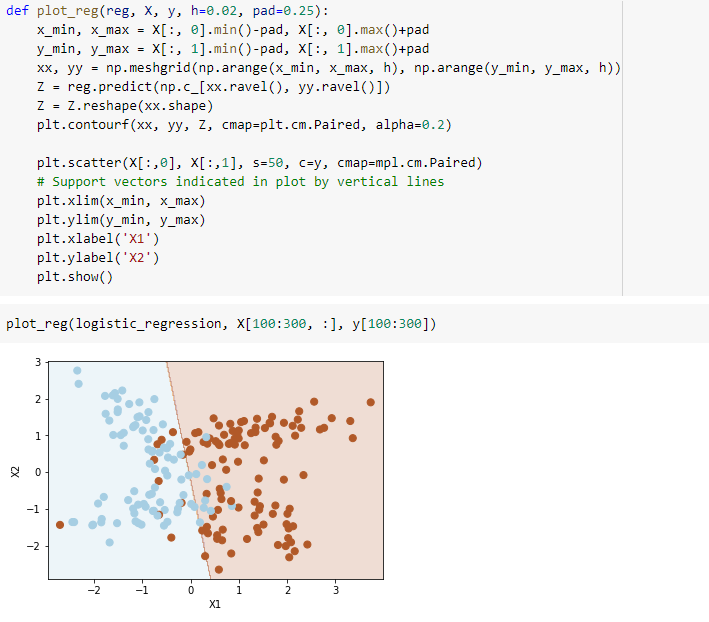
Визуализация SVM:



Логистическая регрессия:



Визуализация Логистической регрессии:



Выводы

В ходе лабораторной работы были исследованы методы логистической регрессии и опорных векторов. Реализованы калибровочные кривые. Было произведено сравнение метода опорных векторов и логистической регрессии.