**Федеральное государственное образовательное**

**бюджетное учреждение**

**высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ**

**ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ**

**ФЕДЕРАЦИИ»**

**(Финансовый университет)**

**Факультет**

**информационных технологий и анализа больших данных**

**Домашнее задание № 6**

«Логистическая регрессия, обучение SVM при помощи прямо-двойственного метода внутренней точки»

Студенты группы ПМ19-2:

Жигулина Юлия

Коротенко Виолетта

Морозов Михаил

Пономаренко Александр

Васильева Александра

Аракелян Рушан

Брашич Илья

Руководитель:

Аксенов Дмитрий Андреевич

Москва 2022

Содержание

1. Алгоритмы
   1. Модель классификации на основе логистической регрессии.
   2. Модель классификации на основе логистической регрессии с радиальными базисными функциями.
   3. Модель классификации на основе логистической регрессии с регуляризацией L1.
   4. Модель классификации на основе метода опорных векторов.

1. Алгоритмы
   1. Модель классификации на основе логистической регрессии.

Входные данные

*–* массив обучающей выборки*, -* вектор признаков для объекта *n, –* его метка класса.

*–* массив предсказываемой переменной.

1. Вычислить значение граничной функции. Обозначим эту величину .
2. Вычислить отношение шансов: .
3. Имея значение , вычислить  с помощью простой зависимости:

4. Получив значение  на первом шаге, на втором, можно записать следующее:

Правая часть уравнения – логистическая функция.

* 1. Модель классификации на основе логистической регрессии с радиальными базисными функциями.

Входные данные:

*–* массив обучающей выборки*, -* вектор признаков для объекта *n, –* его метка класса.

*–* массив предсказываемой переменной.

1. Выбрать несколько точек из пространства состояний. Это позволяет ограничить количество используемых экземпляров.
2. Подставляем признаки в модель линейной регрессии и используем градиентный спуск для обновления параметров модели линейной регрессии.
3. Экземпляры, выбранные вначале, остаются теми же навсегда, признаки не изменяются по мере обучения.
   1. Модель классификации на основе логистической регрессии с регуляризацией L1.

Входные данные

*–* массив обучающей выборки*, -* вектор признаков для объекта *n, –* его метка класса.

*–* массив предсказываемой переменной.

1. Задача оптимизации:
2. Часть компонент оптимального вектора весов , так как решение задачи обладает свойством разреженности.
3. В связи с правилом нулевые веса равносильны исключению соответствующей базисной функции (или исходного признака) из модели.
   1. Модель классификации на основе метода опорных векторов.

Входные данные

*–* массив обучающей выборки*, -* вектор признаков для объекта *n, –* его метка класса.

*–* массив предсказываемой переменной.

1. Метод опорных векторов строит классифицирующую функцию *F* в виде , где  — скалярное произведение, — нормальный вектор к разделяющей гиперплоскости, — вспомогательный параметр.
   1. Если , то объекты попадают в один класс
   2. Если то объекты попадают в другой класс.

*Примечание: Выбор именно такой функции неслучаен: любая гиперплоскость может быть задана в виде для некоторых w и b.*

1. Далее выбираются такие *w* и *b,* которые максимизируют расстояние до каждого класса.

*Примечание: Расстояние равно . Проблема нахождения максимума  эквивалентна проблеме нахождения минимума*.

1. Нахождение максимума производится с помощью множителей Лагранжа.