Московский Авиационный Институт

(Национальный Исследовательский Университет)

Факультет информационных технологий и прикладной математики

Кафедра вычислительной математики и программирования

**Лабораторная работа №1 по курсу**

**«Машинное обучение»**

**Линейные модели**

Студент: Якушкина Мария Константиновна

Группа: М80 – 407Б-19

Дата: 28.10.2022

Оценка: \_\_\_\_\_\_\_\_\_\_\_

Подпись: \_\_\_\_\_\_\_\_\_\_\_

Москва, 2022

* **Постановка задачи**

Вы собрали данные и их проанализировали, визуализировали и представили отчет своим партнерам и спонсорам. Они согласились, что ваша задача имеет перспективу и продемонстрировали заинтересованность в вашем проекте. Самое время реализовать прототип! Вы считаете, что нейронные сети переоценены (просто боитесь признаться, что у вас не хватает ресурсов и данных), и считаете что за машинным обучением классическим будущее и потому собираетесь использовать классические модели. Вашим первым предположением является предположение, что данные и все в этом мире имеет линейную зависимость, ведь не зря же в конце каждой нейронной сети есть линейный слой классификации. В качестве первых моделей вы выбрали, линейную / логистическую регрессию и SVM. Так как вы очень осторожны и боитесь ошибиться, вы хотите реализовать случай, когда все таки мы не делаем никаких предположений о данных, и взяли за основу идею "близкие объекты дают близкий ответ" и идею, что теорема Байеса имеет ранг королевской теоремы. Так как вы не доверяете другим людям, вы хотите реализовать алгоритмы сами с нуля без использования scikit-learn (почти). Вы хотите узнать насколько хорошо ваши модели работают на выбранных вам данных и хотите замерить метрики качества. Ведь вам нужно еще отчитаться спонсорам!

* **Описание**

В данной лабораторной работе реализованы следующие алгоритмы обучения:

1) KNN

идея заключается в определении класса объекта по классам К ближайших(каких

больше - такой и класс).

2) Naive Bayes

остроен на формуле Байеса

3)Линейная регрессия

попытка провести разделяющую гиперилоскость между классами

4)SVM

Линейная с дополнительным условием: максимизируется расстояние от объектов догиперилоскости

* **KNN**

Результаты получились следующие:





Лучший результат при k = 21 и составляет 80% правильных ответов

При использовании sklearn результат практически такой же k = 24

* **Naive Bayes**

1) По гистограммам:





Лучший результат 82% при 22 интервалах

2) Делая предположение, что данные имеют нормальное распределение:





73%

При использовании sklearn: 80%

* **Линейная регрессия**





Результат всего 70%

* **SVM**





Процент угаданного: 72%

* **Вывод**

Данная лабораторная работа дала, мне интересный опыт в работе с настоящими данными. В качестве темы я взяла медицину, так как эта, область, как мне показалось,наиболее хороша с использованием искусственного интеллекта. Реализовав все эти алгоритмы, я получил точность 70-82%, что вполне неплохо. Думаю, что если доработать эту модель с использованием более подходящих алгоритмов и достичь точности около 90%, то ее можно использовать на практике.