

Московский авиационный институт
(национальный исследовательский университет)

Факультет информационных технологий и прикладной
математики

Кафедра вычислительной математики и программирования

Лабораторная работа №2 по курсу «Машинное обучение»

Студент: К. О. Вахрамян
Преподаватель: Ахмед Самир Халид
Группа: М8О-306Б
Дата:
Оценка:
Подпись:

Москва, 2021

Лабораторная работа №2

Задача:

Необходимо реализовать алгоритмы машинного обучения. Применить данные алгоритмы на наборы данных, подготовленных в первой лабораторной работе. Провести анализ полученных моделей, вычислить метрики классификатора. Произвести тюнинг параметров в случае необходимости. Сравнить полученные результаты с моделями реализованными в scikit-learn. Аналогично построить метрики классификации. Показать, что полученные модели не переобучились. Также необходимо сделать выводы о применимости данных моделей к вашей задаче. Задачи со звездочкой бьются по вариантам: N по списку % 2 + 1.

1) ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

2)*SVM - ПЕРВЫЙ ВАРИАНТ

3) ДЕРЕВО РЕШЕНИЙ

4)*RANDOM FOREST - ВТОРОЙ ВАРИАНТ

1 Метод Решения

Для анализа и классификации я взял датасет с информации о годовом доходе физических лиц. Квалификация здесь бинарная: нужно предсказать, доход больше \$50K или нет.

Для анализа результатов я написал функцию `score_prediction`, которая возвращает `accuracy`, `precision`, `recall`, `f1` на основе `TF`, `FP`, `FN`, `TN` из `confusion matrix`.

Гиперпараметры для моих моделей я подбирал, просто проходя по списку из параметров.

2 Выводы

Выполнив данную лабораторную работу, я реализовал различные алгоритмы для классификации данных. Я считаю, что они приемлимы для моей задачи. Наиболее интересным из них мне показался дерево решений. Оно использует интуитивные, понятные идеи, но при этом дает неплохой результат. Однако минус этой модели в том, что она непараметрическая, поэтому ее легко переобучить.