

Лабораторная работа №2

Студента Лунгу Н.А.

Группы М80-304Б-17

Алгоритмы

Логистическая регрессия реализована самым стандартным способом без регуляризаций.

KNN так же вполне стандартен, и реализован с использованием обычной Евклидовой нормы.

При реализации решающего используется критерий разбиения Джини, построение дерева заканчивается при достижении максимальной глубины.

Алгоритм случайного леса реализован на основе алгоритма решающего дерева. При этом используется беггинг и метод случайных подпространств.

Метрики

Для оценки качества классификации использованы метрика ассигасу, так как в обеих задачах отсутствует дисбаланс таргета. Валидирование проводилось посредством кросс-валидации.

Результаты

Алгоритмы логистической регрессии, метода ближайших соседей и решающее дерево в моей и библиотечной реализации показали сравнимые результаты, однако мои версии оказались куда более медленными. Это обусловлено однопоточной реализацией, и отсутствием оптимизаций программы. Если с результатами линейной модели и KNN еще можно смириться, то с деревом дела обстоят куда хуже.

Библиотечная реализация случайного леса превосходит самодельную по качеству на обоих датасетах.

Не стал разбираться в теме переобучения из-за того, что уже октябрь месяц и надо бы делать третью лабораторную работу по ИИ.

Возникшие проблемы

Основные трудности, как и полагается, состояли в понимании алгоритма и последующей его реализации. Из представленных в ноутбуке результатов заметно, что некоторые реализованные алгоритмы себя показали намного медленнее библиотечных версий, что также связано с некоторыми костылями в коде.

Выводы

В результате работы над реализацией алгоритмов машинного обучения стало понятно, что делать это стоит исключительно в образовательных целях. Не стоит "изобретать

велосипед", когда существуют общедоступные реализации алгоритмов, значительно превосходящие самодельные.