

Курс по методам машинного обучения

Практическое задание № 3

KNN

Драгунов Никита

1 Описание задания

В настоящем задании вы познакомитесь с алгоритмом k ближайших соседей (k -NN, KNN, k NN) для решения задач классификации и регрессии.

Приведем здесь краткое напоминание о принципе работы k NN. Пусть дана обучающая выборка $X = (x_i, y_i)$ и функция расстояния ρ . Требуется классифицировать новый объект u . Алгоритм k ближайших соседей относит объект u к тому классу, представителей которого окажется больше всего среди k его ближайших по ρ соседей: $\alpha(u; X, k) = \underset{y \in Y}{\operatorname{argmax}} \sum_{i=1}^k w_i [y_u^{(i)} = y]$, где $y_u^{(i)}$ — метка класса i -го соседа объекта u . В классическом методе k ближайших соседей все объекты имеют единичные веса: $w_i = 1$. Альтернативой данному подходу служат веса, обратно пропорциональные расстоянию между объектами. При решении задачи регрессии ответом алгоритма служит средневзвешенное значение меток $y_u^{(i)}$ среди k ближайших соседей.

2 Формат задания

Задание состоит из двух различных видов подзаданий: **кросс-рецензирование** и **юнит-тестирование**. В этом задании за юнит-тестирование можно получить максимум **15 баллов**, а за кросс-рецензирование — **25 баллов**.

Юнит-тестирование — уже знакомый вам формат, в котором необходимо реализовать какие-либо функции. В данном задании вам необходимо реализовать функции, находящиеся в файлах `scalers.py` и `cross_val.py` (в ходе выполнения заданий 1.1 и 2.1 из ноутбука). Их можно найти в шаблоне решения во вкладке KNN (`unit-tests`). После реализации ваш код можно протестировать локально, а затем его необходимо сдать в проверяющую систему (**вкладка KNN (`unit-tests`)**). Напомним, что такой вид заданий можно сдавать после дедлайна со штрафом 40%.

Кросс-рецензирование — в таких заданиях вам необходимо написать код, ответить на вопросы, что-то визуализировать в ноутбуке KNN.ipynb, который можно найти в шаблоне решения во вкладке KNN (ноутбук). Выполненный ноутбук KNN.ipynb необходимо **сдать в тестирующую систему во вкладку KNN (ноутбук)**. Данный вид задания принимается **только до дедлайна**. После окончания срока сдачи, у вас будет еще **неделя** на проверку решений как минимум **3х других студентов** — это **необходимое** условие для получения оценки за вашу работу. Если вы считаете, что вас оценили неправильно, можете писать на почту `ml.cmc@mail.ru` с темой письма `ВМК.ML[Задание 3][peer-review]` с просьбой перепроверить оценивание задания.

Важно: перед сдачей проверьте, пожалуйста, что не оставили в ноутбуке где-либо свои ФИО, группу и т.д. — кросс-рецензирование проводится анонимно.

Важно: запрещается пользоваться библиотеками, импорт которых не объявлен в файлах с шаблонами функций.

Важно: задания, в которых есть решения, содержащие в каком-либо виде взлом тестов, дополнительные импорты и прочие нечестные приемы, будут автоматически оценены в 0 баллов без права пересдачи задания.

Важно: перед сдачей задания на кросс-проверку не забудьте перезапустить ноутбук: *Kernel -> Restart & Run All*.

3 Тестирование

Реализованные модули предлагается проверить с помощью юнит-тестов, которые расположены в `public_tests.zip`. Разархивируйте архив с тестами в директорию `tests`, положите рядом файл `run.py` и реализованные вами модули. Добавьте ему права на выполнение скрипта `run.py` (`chmod +x run.py` в консоли). Запустите тесты `./run.py unittest scalars` и `./run.py unittest cv`.