В первой части задания вы будете настраивать линейную модель для прогнозирования количества прокатов велосипедов в зависимости от календарных характеристик дня и погодных условий. Нужно так подобрать веса признаков, чтобы уловить все линейные зависимости в данных и в то же время не учесть лишние признаки, тогда модель не переобучится и будет делать достаточно точные предсказания на новых данных. Найденные линейные зависимости нужно будет интерпретировать, то есть понять, соответствует ли обнаруженная закономерность здравому смыслу. Основная цель задания – на примере показать и объяснить, из-за чего возникает переобучение и как с ним можно бороться.

Для выполнения задания используйте ipython notebook:

OverfittingTask.ipynb

Нужно написать код и кратко ответить на вопросы. Подробные инструкции даны в начале notebook.

Данные из репозитория UCI – информация об автоматизированном прокате велосипедов:

bikes\_rent.csv