

## Задание на нейронные сети.

В репозитории UCI:

- (1) выбрать дейтасет,
- (2) посмотреть его,
- (3) найти место целевого столбца (обычно это первый или последний столбец),
- (4) определить количество меток классов (class labels) в целевом столбце,
- (5) переставить местами строки (с помощью скрипта Python и функций библиотеки pandas) так, чтобы данные одного класса были расположены подряд,
- (6) произвести дискретизацию числовых столбцов (один числовой столбец превращается при этом в несколько – обычно три – категориальных) (функция cut),
- (7) произвести бинаризацию категориальных столбцов (функция get\_dummies),
- (8) для однослойного персептрона перевести данные из 0, 1 в -1, 1,
- (9) выделить данные для обучения (80%) и для тестирования (20%),
- (10) выделить входные и желаемые выходные данные (желаемые выходные данные получаются из столбцов, в которые превратился целевой столбец – этих столбцов будет столько, сколько меток классов было в целевом столбце), это выделение происходит как в данных для обучения, так и в данных для тестирования,
- (11) применить однослойный персептрон,
- (12) применить многослойный персептрон,
- (13) записать результаты тестирования,
- (14) подготовить краткий отчет: какой дейтасет использовался, сколько столбцов и сколько строк в дейтасете, имена столбцов, какой по порядку целевой столбец, каковы результаты тестирования для однослойного персептрона, каковы результаты тестирования для многослойной сети (многослойного персептрона) – (каково количество и процент ошибок при тестировании).

Пункты (6) и (8) могут не выполняться при подготовке данных для многослойного персептрона.

Рекомендация: проверьте, как выполнить все эти требования для дейтасета про съедобные и ядовитые грибы.

Репозиторий UCI и дейтасет mushrooms в нем найти через поисковые системы.