

## Практика № 5. Вариант 1

Необходимо в зависимости от варианта сгенерировать датасет и сохранить его в формате csv.

Построить модель, которая будет содержать в себе автокодировщик и регрессионную модель.

В качестве результата представить исходный код, сгенерированные данные в формате csv, кодированные и декодированные данные в формате csv, результат регрессии в формате csv (что должно быть и что выдает модель), и сами 3 модели в формате h5.

Исходные данные:

$X \sim N(3, 10)$

$e \sim N(0, 0.3)$

Признак	1	2	3	4	5	6	6
Формула	$X^2 + e$	$\sin(X/2) + e$	$\cos(2 \cdot X) + e$	$X - 3 + e$	$-X + e$	$ X  + e$	$(X^3)/4 + e$

## Реализация

В функции *genData* генерируется датасет с помощью функций из стат пакета *scupy.stats*. В функции *splitData* набор данных разделяется на три части – тренировочный, валидационный и тестовый. Основная функция программы – *createModel*, в которой происходит создание всех трёх моделей. Каждая из моделей представляет собой неглубокую нс – один или два скрытых слоя с функцией активаций Relu, батч нормализацией и дропаутом.

Сначала обучается одна нс, выполняющая функции всех трёх. Но после обучения часть весов общей сети передаётся всем “маленьким” – кодировщику, декодировщику и регрессии. Каждая из моделей сохраняется в соответствующий файл расширения h5. Также создаются csv файлы с кодированными и декодированными тестовыми данными, а также файл с результатом работы регрессии.