

## Домашнее задание по машинному обучению №5

Дедлайн на полный балл – 19 апреля

Дедлайн на половинный балл – 26 апреля

Датасеты – `spambase_old.csv`, `spambase_new.csv`, `mnist.csv`

0. В задачах 2, 3, 5 градиентный спуск нужно делать пока уменьшается ошибка на val датасете.

1. **(2)** Обратное распространение градиента.

Напоминание — для последнего слоя нейронной сети  $\delta$  выражается как:

$$\delta_i^L = \frac{\partial L(w)}{\partial s_i^L}$$

Выпишите в явном виде  $\delta^L$  для softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь.

2. **(5)** Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете `spambase_old` (train) и проверьте на датасете `spambase_new` (val). Получите ROC кривые для вариантов без нормировки и с нормировкой признаков.
3. **(6)** Модифицируйте модель из задачи 2, добавив 2 скрытых слоя (по 64 нейрона) с функциями активации ReLU в них и реализовав алгоритм обратного распространения градиента.
4. **(1)** Разделите датасет `mnist.csv` на train (80%), val (10%), test (10%) с сохранением пропорций классов. Убедитесь, что пропорции сохранились.
5. **(4)** Модифицируйте модель из задачи 3, заменив последний нейрон на 10 нейронов, и реализовав мультиклассовую классификацию с softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь.

## Задачи на дополнительные баллы

Дедлайн на полный балл – 26 апреля

Дедлайн на половинный балл – 3 мая

1. **(6)** Решите задачу 5 с помощью kNN, дерева и случайных лесов (проведя разумный подбор параметров по val датасету). Сравните результаты на test датасете.