

1 (1 балл) Производные

Вычислите производную по $x \in \mathbb{R}^n$ для следующих функций:

- (a) $f(x) = x$;
- (b) $g(x) = x^\top x$. Подсказка: сначала найдите $x^\top x$.

2 (1 балл) Функции активации

Вычислите производную функции $\tanh(z)$ по z . Изобразите графики функции и ее производной. Покажите, что при больших по модулю значениях z градиенты затухают.

3 (3 балла) Функции потерь

В задаче (многоклассовой) классификации на K классов для обучения используется кросс-энтропия. Общая формула для нее выглядит следующим образом (в случае одного объекта):

$$CE(w) = - \sum_{k=1}^K p_k \log \hat{y}_k.$$

p_k здесь обозначает истинную вероятность принадлежности объекта классу k . Во многих задачах вектор вероятностей p является one-hot-вектором, т. е. имеет вид $[0, \dots, 0, 1, 0, \dots, 0]$, поэтому работает формула из лекции 1. Однако бывает и по-другому: например, из четырех разметчиков MNIST трое выбрали класс 1, а один — класс 7.

Тогда можно положить $p = [0, \frac{3}{4}, 0, 0, 0, 0, \frac{1}{4}, 0, 0]$.

Пусть $\hat{y} = \text{softmax}(z)$.

- 1. (2 балла) Найдите производную функции потерь по вектору z в случае, когда целевая переменная принадлежит ровно к одному классу. Подсказка: воспользуйтесь формулой из лекции 1. Считайте, что размер батча равен 1.
- 2. (1 балл) Прodelайте то же самое для произвольного вектора вероятностей p .

4 (1 балл) Полносвязные нейронные сети

На вход поступает вектор x размерности 3, далее он проходит через три полносвязных слоя с размерами выходов 5, 2 и 4 соответственно. Выпишите формулу для вычисления результата нейронной сети (смещение присутствует на каждом слое). Нарисуйте схему такой сети.

5 (1 балл) Оптимизаторы

Изучите формулы обновления весов для оптимизаторов SGDW и AdamW (<https://arxiv.org/pdf/1711.05101.pdf>, с. 3). Перепишите их в обозначениях из лекции 2.

6 (1 балл) Инициализация

Объясните смысл параметра `gain` в PyTorch (<https://pytorch.org/docs/stable/nn.init.html>). Что произойдет, если выбрать неправильное значение этого параметра?

7 (1 балл) Батч-нормализация

Сверточный слой принимает на вход батч размерности $B \times h \times w \times C$. Какие параметры добавится, если перед сверточным слоем поставить батч-нормализацию? Сколько будет таких параметров?

8 (1 балл) Обучение нейронных сетей

Объясните идею следующих техник.

1. Ранняя остановка (early stopping).
2. Переменная скорость обучения (LR scheduler).
3. Регуляризация.
4. Дропаут.