# Машинное обучение, ФКН ВШЭ
# Домашнее задание №9

**Задача 1.** На семинаре при выводе метода обратного распространения ошибки мы использовали формулу производной сложной функции, но при этом все наши переменные были скалярами. В реальных алгоритмах приходится работать с векторами, и для этого потребуется формула сложной производной в векторном виде.

Пусть $x \in \mathbb{R}^n, y \in \mathbb{R}^m, z \in \mathbb{R}$, и заданы функции

$$z = g(f(x)), \quad y = f(x). \tag{0.1}$$

Выведите формулу производной сложной функции в векторном виде.

Теперь мы можем повторить все рассуждения из семинара, но уже для векторов. Возникает вопрос — что делать с более сложными структурами весов? Ведь в общем случае мы представляем параметры и выходы очередного слоя в виде тензоров. На самом деле, запись в виде тензора используется для более удобного представления формул, таким образом, мы можем векторизовать все наши формулы и использовать формулу сложной производной для векторов.

Следующие пункты посвящены вопросу сходимости метода градиентного спуска, который мы так активно используем для оптимизации функционалов качества.

Чтобы показать сходимость градиентного спуска, нам потребуются неравенства из задач 2, 3, 4.

**Задача 2.** Рассмотрим $x \in \mathbb{R}^n$ и функцию $f(x)$. Пусть градиент $f(x)$ – липшицев в $L_2$ с константой $L$.
Докажите неравенство

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leqslant \frac{L}{2}\|y - x\|^2. \tag{0.2}$$

**Задача 3.** Пусть $x \in \mathbb{R}^n$, функция $f(x)$ выпукла, а градиент функции $f(x)$ – липшицев в $L_2$ с константой $L$.
Докажите неравенство

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leqslant f(y). \tag{0.3}$$

**Решение.**

Рассмотрим функцию

$$\varphi(y) = f(y) - \langle \nabla f(x_0), y \rangle, \tag{0.4}$$

$$\nabla \varphi(y) = \nabla f(y) - \nabla f(x_0), \tag{0.5}$$

где $x_0$ – некоторая фиксированная точка. Легко проверяется, что функция дифференцируема, выпукла, а градиент – липшицев с константой $L$.

Используя результат предыдущей задачи, для выпуклой функции можем записать

$$\varphi(y) - \varphi(x) - \langle \nabla\varphi(x), y - x\rangle \leqslant L\|y - x\|^2. \tag{0.6}$$

Выберем в качестве $y = x - \frac{1}{L}\nabla\varphi(x)$, тогда нер-во переписывается в следующем виде:

$$\varphi\left(x - \frac{1}{L}\nabla\varphi(x)\right) - \varphi(x) + \frac{1}{L}\|\nabla\varphi(x)\|^2 \leqslant \frac{1}{2L}\|\nabla\varphi(x)\|^2 \tag{0.7}$$

$$\varphi(x_0) \leqslant \varphi\left(x - \frac{1}{L}\nabla\varphi(x)\right) \leqslant \varphi(x) - \frac{1}{2L}\|\nabla\varphi(x)\|^2 \tag{0.8}$$

$$f(x_0) - \langle \nabla f(x_0), x_0\rangle \leqslant f(x) - \langle \nabla f(x_0), x\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(x_0)\|^2 \tag{0.9}$$

$$f(x_0) + \langle \nabla f(x_0), x - x_0\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(x_0)\|^2 \leqslant f(x) \tag{0.10}$$

■

**Задача 4.** Пусть $x \in \mathbb{R}^n$, функция $f(x)$ выпукла, а градиент функции $f(x)$ липшицев в $L_2$ с константой $L$.
Докажите неравенство

$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leqslant \left\langle \nabla f(x) - \nabla f(y), x - y\right\rangle. \tag{0.11}$$

После того, как мы вывели основные неравенства, можем приступить к доказательству сходимости градиентного спуска.

**Задача 5.** Пусть $x \in \mathbb{R}^n$, функция $f(x)$ выпукла, а градиент функции $f(x)$ липшицев в $L_2$ с константой $L$.
Рассмотрим метод градиентного спуска

$$x_{k+1} = x_k - \alpha\nabla f(x_k) \tag{0.12}$$

Покажите что он сходится при шаге $\alpha < \frac{2}{L}$.

**Решение.**

Обозначим $d_k = \|x_k - x^*\|$, где $x_k$ – позиция на шаге $k$, $x^*$ – точка оптимума. Покажем, что на шаге $k+1$ мы будем ближе к точке оптимума, чем на шаге $k$.

$$d_{k+1}^2 = \|x_k - \alpha\nabla f(x_k) - x^*\|^2 = \tag{0.13}$$

$$= d_k^2 - 2\alpha\left\langle \nabla f(x_k), (x_k - x^*)\right\rangle + \alpha^2\|\nabla f(x_k)\|^2 \leqslant \ldots \tag{0.14}$$

Воспользуемся неравенством из задачи 4 и тем, что $\nabla f(x^*) = 0$

$$\ldots \leqslant d_k^2 + \alpha^2\|\nabla f(x_k)\|^2 - \frac{2\alpha}{L}\|\nabla f(x_k)\|^2 = d_k^2 - \alpha\left(\frac{2}{L} - \alpha\right)\|\nabla f(x_k)\|^2 \qquad (0.15)$$

Получаем, что $d_{k+1} \leqslant d_k$, т.е. на каждом шаге мы приближаемся к точке минимума.

Воспользуемся неравенством из задачи 2 и выпуклостью функции $f(x)$ и запишем:

$$f(x_{k+1}) \leqslant f(x_k) + \langle\nabla f(x_k), x_{k+1} - x_k\langle + \frac{L}{2}\|x_k - x_{k+1}\|^2 \leqslant f(x_k) - \left(\alpha - \frac{\alpha^2 L}{2}\right)\|\nabla f(x_k)\|^2 \qquad (0.16)$$

Обозначим $\delta_k = f(x_k) - f(x^*) > 0$, тогда можем из обеих частей неравенства вычесть $f(x^*)$ и переписать его в виде:

$$\delta_{k+1} \leqslant \delta_k - \left(\alpha - \frac{\alpha^2 L}{2}\right)\|\nabla f(x_k)\|^2 \qquad (0.17)$$

Теперь необходимо ограничить снизу норму градиента – это можно сделать, записав условие выпуклости функции

$$\delta_k = f(x_k) - f(x^*) \leqslant \langle\nabla f(x_k), x_k - x^*\rangle \leqslant d_k\|\nabla f(x_k)\| \leqslant d_0\|\nabla f(x_k)\| \qquad (0.18)$$

Тогда наше неравенство можно переписать в виде:

$$\delta_{k+1} \leqslant \delta_k - \left(\alpha - \frac{\alpha^2 L}{2}\right)\|\nabla f(x_k)\|^2 \leqslant \delta_k - \left(\alpha - \frac{\alpha^2 L}{2}\right)\frac{\delta_k^2}{d_0^2} \qquad (0.19)$$

Разделим неравенство на $\delta_k\delta_{k+1}$:

$$\frac{1}{\delta_k} \leqslant \frac{1}{\delta_{k+1}} - \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right)\frac{\delta_k}{\delta_{k+1}} \qquad (0.20)$$

$$\frac{1}{\delta_{k+1}} \geqslant \frac{1}{\delta_k} + \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right)\frac{\delta_k}{\delta_{k+1}} \geqslant \frac{1}{\delta_k} + \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right) \qquad (0.21)$$

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} \geqslant \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right) \qquad (0.22)$$

Просуммируем эти неравенства и получим:

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_0} \geqslant \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right)(k+1) \qquad (0.23)$$

$$\delta_{k+1} \leqslant \frac{1}{\frac{1}{\delta_0} + \frac{1}{d_0^2}\left(\alpha - \frac{\alpha^2 L}{2}\right)(k+1)} \tag{0.24}$$

Из последнего неравенства следует, что

$$\lim_{k\to\infty} \delta_{k+1} = 0 \tag{0.25}$$

■