# Глубинное обучение 1, ФКН ВШЭ
# Теоретическое домашнее задание №1
# Полносвязные нейронные сети

**Задача 1.** Ниже приведен код вычислений на PyTorch. Нарисуйте граф вычислений, который реализует этот код. Затем сделайте проход вперед и проход назад по графу, возле каждой вершины подпишите два значения, которые соответствуют ей при проходе вперед и назад (можно рисовать граф поэлементно). Что будет записано в полях a.grad, b.grad, c.grad после выполнения этого фрагмента кода?

```
a = torch.tensor([1.0, 1.0], requires_grad=True)
b = torch.tensor([1.0, -1.0], requires_grad=False)
c = torch.tensor([-1.0, 2.0], requires_grad=True)

l = torch.relu(a * b).sum() + (a + c ** 2).prod()
l.backward()
```

**Задача 2.** Пусть $x \in \mathbb{R}^{d_1}, W_1 \in \mathbb{R}^{d_2 \times d_1}, W_2 \in \mathbb{R}^{d_3 \times d_2}, W_3 \in \mathbb{R}^{d_3}$ и $y \in \mathbb{R}$. Определим $\ell$ как:

$$\ell = \Big(W_3^T \cos\big(W_2 \sin(W_1 x)\big) - y\Big)^2,$$

где тригонометрические функции берутся поэлементно. Подсчитайте градиенты $\ell$ по $W_1, W_2$ и $W_3$, запишите их в матрично-векторном виде.

**Задача 3.** Рассмотрим задачу многоклассовой классификации с $K$ равновероятными классами. Пусть логиты предсказания $z_k, k = 1, \dots, K$ генерируются независимо от целевой переменной $y$ из распределения с нулевым средним и конечным экспоненциальным моментом: $\mathbb{E}[e^z] < \infty$. Докажите, что матожидание кросс-энтропии в таком случае ограничено сверху логарифмом числа классов плюс некоторая константа:

$$\mathbb{E}\Big[\mathcal{L}_{\text{CE}}(y, z)\Big] \le \log(K) + C$$

Найдите эти константу для случая нормальных логитов $z_k \sim \mathcal{N}(0, \sigma^2)$.

*Определение:* Функция $g : X \to \mathbb{R}$ равномерно приближает функцию $f : X \to \mathbb{R}$ на множестве $X$ с ошибкой $\varepsilon > 0$, если:

$$\sup_{x \in X} \Big| f(x) - g(x) \Big| \le \varepsilon$$

**Задача 4.** Рассмотрим класс функций $G_\sigma$, которые задаются полносвязными нейронными сетями с конечной шириной и глубиной, с функцией активации $\sigma$. Можно ли равномерно приблизить произвольную непрерывную функцию $f$ на множестве $\mathbb{R}$ с помощью функции $g \in G_\sigma$ с произвольной наперед заданной ошибкой $\varepsilon > 0$, если:

(a) $\sigma(x) = \mathrm{ReLU}(x)$?

(b) $\sigma(x)$ — произвольный многочлен?

**Задача 5.** Докажите теорему Цыбенко для одномерного признакого пространства и функции активации ReLU: любую функцию $f(x)$, непрерывную на отрезке $[a, b]$ можно равномерно приблизить конечной полносвязной нейронной сетью с одним скрытым слоем и активацией ReLU с любой наперед заданной ошибкой $\varepsilon > 0$.

*Подсказка:*

(a) Докажите, что любую непрерывную на отрезке функцию можно равномерно приблизить кусочно-линейной функцией.

(b) Покажите, как произвольную кусочно-линейную функцию выразить с помощью полносвязной нейросети с одним скрытым слоем.