Theoretical assignment 3; 10 points total + 5 points extra

Theoretical Deep Learning course, MIPT

Problem 1

4 points total.

A Gram matrix G of a set of vectors $\{v_i\}_{i=1}^n$ from Euclidean space V with scalar product $\langle \cdot, \cdot \rangle$ is defined as $G_{ij} = \langle v_i, v_j \rangle$.

1. **1 point.** Prove that Gram matrix is positive semi-definite (PSD).
2. **2 points.** Prove that H^∞ (see lecture slides) is indeed a Gram matrix.
3. **1 point.** Prove that $H(W)$ is a Gram matrix too.

Problem 2

2 points.

Consider the same setting as in Du et al. (2018)[1], but with sigmoid instead of ReLU:

$$f(W, a, x) = \frac{1}{\sqrt{m}} \sum_{r=1}^{m} a_r \sigma(w_r^T x),$$

where $x \in \mathbb{R}^d$, $w_r \in \mathbb{R}^d$, $a_r \in \mathbb{R}$ and $\sigma(z) = \log(1 + \exp(z))$. We consider l_2-regression problem:

$$\mathcal{L}(W) = \frac{1}{2} \sum_{i=1}^{n} (f(W, a, x_i) - y_i)^2.$$

Parameters are initialized as follows:

$$w_r \sim \mathcal{N}(0, I), \ a_r \sim U(\{-1, 1\}) \quad \forall r = 1, \ldots, m.$$

Data points are assumed to be normalized: $\|x_i\|_2 = 1, \ |y_i| < C \quad \forall i = 1, \ldots, n.$

Prove Lemma 3.1 of Du et al. (2018) in this setting (with the same, up to a constant, lower bound for m).

We expect you to present a more detailed proof compared to one given in the paper.

[1] `https://openreview.net/forum?id=S1eK3i09YQ`

Problem 3

4 points + 5 points extra.

Prove Lemma 3.2 of Du et al. (2018) in this setting (with the same, up to a constant, lower bound for R).

We will give you up to 5 extra points for proving better bounds for R.