Theoretical Foundations for Deep Learning: Problem Set # 1

Instructor: Ankur Moitra

Due: March 24th

You can work with other students, but you must write-up your solutions by yourself and indicate at the top who you worked with!

Please submit by March 24th, at 11:59pm. I have set up a course on gradescope and you should be able to add the course with the access code 86D8KV. If you have any trouble, please email me directly.

Problem 1 30 points

Consider the parity function $\chi : \{0,1\}^n \to \{0,1\}$.

(a) Construct a deep network with ReLU activations that exactly expresses the parity function with $O(n)$ hidden units.

(b) Consider a ReLU network with m hidden units. Show that the function computed by the network can be described by a partition of the space of its inputs into at most $f(m)$ parts, where the function is linear on each one. What is your $f(m)$?

(c) ~~Use part (b) above to show that there is no deep network that expresses the parity with $o(n)$ hidden units.~~ There was a bug! Skip this problem

Problem 2 30 points

We say that a function $f : \mathbb{R} \to \mathbb{R}$ is positive semidefinite if for all sequences $x_1, x_2, \cdots, x_k$ we have that $\sum_{i,j} x_i x_j f(x_i - x_j) \geq 0$. Bochner's Theorem gives us an equivalent characterization that a function is positive semidefinite iff it can be written as

$$f(x) = \int_\omega e^{2\pi i \omega^T x} F(\omega) d\omega$$

for some nonnegative function F. Show that

$$C_f \leq O\Big(-f(0)\nabla^2 f(0)\Big)$$

where C_f is the constant in Barron's Theorem. Note that positive semidefinite functions have $\nabla^2 f(0) \leq 0$. *Hint: Use Cauchy-Schwarz*

Problem 3 20 points

Consider a training set $x_1, x_2, \cdots, x_N$. Assuming that all the x_i's are distinct, show that there is always a function $\phi : \mathbb{R}^n \to \mathbb{R}^N$ so that for any with labels $y_1, y_2, \cdots, y_N \in \{+1, -1\}$ there is a halfspace that perfectly labels the training set. In particular $\text{sgn}(a^T\phi(x_i) + b) = y_i$ for all i. *Note: neural tangent kernels would give such a mapping but for a target dimension that is polynomial in N*

Problem 4 20 points

Define the reparametrization $x = g(u)$ where g acts element-wise and $g(u_i) = (u_i)^{2\ell}$ for some positive integer ℓ. Consider the gradient flow on the objective

$$f(u) = \frac{\|Ag(u) - b\|_2^2}{2}$$

Find an ODE describing the trajectory of x.