# CSE 599 Theoretical Deep Learning Homework 1

February 6, 2021

- Deadline: Feb. 15th. No late homework.
- Include your name and UW NetID in your submission.
- Homework must be typed. You can use any typesetting software you wish (latex, markdown, ms word, etc).
- You may discuss assignments with others, but you must write down the solutions by yourself.

## 1 Univariate Approximation With Shallow ReLU Networks (10 Points)

Let $g : [0,1] \to \mathbb{R}$ be an arbitrary twice-differentiable function with $g(0) = g'(0) = 0$.

1. (3 Points) Prove the infinitely-wide representation $g(x) = \int_0^1 \sigma(x-b) g''(b) db$.
2. (3 Points) Suppose $|g''| \leq \beta$ over $[0,1]$ for some $\beta > 0$, and let $\epsilon > 0$ be given. Prove that there exists a ReLU network $f(x) \triangleq \sum_{i=1}^m a_i \sigma(x - b_i)$ with $m \leq \left\lceil \frac{\beta}{\epsilon} \right\rceil$ and $\|f - g\|_\infty \leq \epsilon$.
3. (4 Points) Use 1.2 and the Sampling Lemma from Jan. 13th's lecture (Pister's Lemma) to prove that for any $\epsilon > 0$, there exists a ReLU network $f(x) \triangleq \sum_{i=1}^m a_i \sigma\left(x - b_i\right)$ with

$$m \leq \left\lceil \frac{1}{\epsilon} \left[ \int_0^1 \left|g''(x)\right| \right]^2 \right\rceil \text{ and } \int_0^1 \left(f(x) - g(x)\right)^2 dx \leq \epsilon.$$

## 2 Gradient Descent Learns Single-Neuron Neural Network with Gaussian Input (10 Points)

Consider the following optimization problem:

$$\min_{w \in \mathbb{R}^2} f(w) = \mathbb{E}_{x \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)} \left[ \left( \sigma\left(x^\top w\right) - \sigma\left(x^\top w^*\right) \right)^2 \right]$$

where $\sigma(\cdot)$ is the ReLU activation function, $\mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ is the standard 2-dimensional Gaussian distribution with zero mean and an identity covariance, and $w^* \neq 0$ is an unknown vector. Consider the gradient descent dynamics starting from some initial point $w_0$ with $\theta(w_0, w^*) \neq \pi$ and $0 \neq \|w_0\|_2 \leq \|w^*\|_2$

$$w_{t+1} = w_t - \eta \nabla f(w_t) \text{ for } t = 1, 2, \ldots \tag{1}$$

$\theta(\cdot,\cdot)$ is the angle between two vectors, and is within $[0, \pi]$.

1. (4 points) Show that for $w \neq 0$

$$\begin{aligned} f(w) =& \frac{1}{2}\|w\|_2^2 - \frac{1}{\pi}\|w\|_2 \|w^*\|_2 \left(\sin\theta(w, w^*) + (\pi - \theta(w, w^*))\cos\theta(w, w^*)\right) + \frac{1}{2}\|w^*\|_2^2, \\ \nabla f(w) =& w - \frac{1}{\pi}\frac{\|w^*\|_2 \sin\theta(w, w^*)}{\|w\|_2} w - \frac{1}{\pi}(\pi - \theta(w, w^*)) w^* \end{aligned}$$

2. (2 points) Show the only differentiable critical point of $f(w)$ is $w^*$ (you can ignore the case where $w = 0$).

3. (2 points) Prove that for $\eta < 1$, $\theta(w_t, w^*) \geq \theta(w_{t+1}, w^*)$.

4. (2 points) Show if the initialization $w_0 \neq 0$ and $\theta(w_0, w^*) < \pi$, then for small $\eta$, $w_t$ will NOT converge to 0. (Hint: Show $\|w_t\|_2$ is bounded away from 0.)

5. (Bonus: no point) Prove if the initialization $w_0 \neq 0$ and $\theta(w_0, w^*) < \pi$, then for sufficiently small $\eta$ (which can depend on $\|w^*\|_2$ and $w_0$), $w_t$ will converge to $w^*$.

## 3 Landscape of Top Eigenvector Problem (5 Points)

Consider the following optimization problem for computing the top eigenvector of a positive semi-definite matrix $M \in \mathbb{R}^{d\times d}$:

$$\min_{x\in\mathbb{R}^d} f(x) = \frac{1}{4}\left\|xx^\top - M\right\|_F^2.$$

Let the eigen-decomposition of $M$ be $M = \sum_{i=1}^d \lambda_i v_i v_i^\top$ where $\{\lambda_i\}_{i=1}^d$ are eigenvalues and $\{v_i\}_{i=1}^d$ are eigenvectors. Assume $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \lambda_d \geq 0$. Prove that all saddle points and local maxima are strict, i.e., except the global minima, all other first-order stationary points' Hessian have a negative eigenvalue.