A Hardware

All MKL experiments were conducted on Intel® Xeon® E5-2650 v2 CPUs, 2.60GHz with 64GB RAM and eight cores. All CKL and CNN experiments were conducted on a cluster with 32 nodes, each node has Intel® Xeon® E5-2660 CPUs, 2.20GHz with 64GB RAM and eight cores, and two Nvidia Tesla K20m GPUs.

B Proofs

Lemma 3. *The decision function* $\mathbf{1}_{\mathrm{Im}(z(wx+\beta))\geq 0}$ *induces a unique binary labeling for the set* $x \in \{1/2^i\}_{i=1}^n$ *for every integer value of* $w \in [1..2^n]$, *and any* $\beta \in (0, 2^{-(n+1)})$.

Proof. For any integer $w \in 1..2^n$ and $i \in 1..n$, choose the binary label as 0 if $z(w/2^i + \beta)$ lands in the upper half-plane of $\mathbb{C}$, and 1 if the lower half-plane. The label can be read as the most significant fractional digit of the binary representation of $w/2^i$, as long as $\beta \in (0, 2^{-(n+1)})$[5]. The labeling is then unique for integer values of w up to 2^n. □

An example of the proof construction can be seen in Figure 1, for $n = 2$.

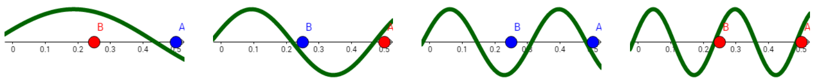

Fig. 1. An example of Lemma 3, with $n = 2$. The values of w increase from 1 to 4, proceeding left to right. The positive label is represented as a blue point, while negative is red.

Lemma 4. *The shatter function of* $(\mathbb{R}^d, \mathcal{G}_d(R))$ *is* $O(R^d n^{d+1})$.

Proof. We can first observe that $\|\omega\|_2 \leq R$ implies that $\|\omega\|_\infty \leq R$. This implies that $|\omega_j| \leq R$ for every $j \in [1..d]$. Treating each coordinate separately this way, each term in $\langle \omega, \mathbf{x} \rangle + \beta$ contributes a factor in the growth function.

For a fixed ω, the number of subsets of a set of n points selected by (ω, β, d)-ranges is $O(n)$, because as β changes, at most one point exits or leaves the upper half-plane (because the points all travel at the same speed around the unit circle).

For fixed β, and fixed ω save for some coordinate ω_j, on the other hand, how often a point enters or leaves the upper half-plane as ω_j varies in $(0, R]$ depends upon the value of x_j. For higher values of x_j, the mapped point travels more rapidly. In fact, for $x = 1$, z takes R revolutions around the circle, so enters and exits the upper half-plane $2R$ times. The number of subsets is bounded by

$$\sum_{i=1}^{n} 2R|x_i| = 2R \sum_{i=1}^{n} |x_i| \leq 2Rn. \tag{6}$$

[5] To avoid ambiguity, we require $\beta > 0$, to prevent $z(w/2^i)$ from landing on the real axis when 2^i divides w.

We take the absolute value because a negative x_i simply changes the direction of travel of $z(\omega_j x_i + \beta)$. Everything else remains the same. For ω and β varying independently, we now have the bound stated in the lemma. $\square$