CSE 251A: Homework 5 Solutions

1. *Monotone disjunctions.*

   (a) Any subset I of features $\{1, 2, \ldots, d\}$ corresponds to a disjunction $\vee_{i \in I} x_i$, and every disjunction in $\mathcal{H}$ corresponds to a subset of features $\{1, 2, \ldots, d\}$. Hence, there are as many disjunctions as there are subsets of features, so $|\mathcal{H}| = 2^d$.

   (b) Using the PAC bounds for the separable case that we studied in class, the true error of h can be bounded thus, with probability at least $1 - \delta$:

   $$\text{err}(h) \leq \frac{1}{n} \ln \frac{|\mathcal{H}|}{\delta} = \frac{1}{n} \left(d \ln 2 + \ln \frac{1}{\delta} \right).$$

   (c) The set of all k-sparse disjunctions again correspond to all subsets of $\{1, 2, \ldots, d\}$ of size k or less. The number of such subsets is

   $$N = \binom{d}{1} + \binom{d}{2} + \ldots + \binom{d}{k}$$

   (Very) approximately, this is at most $2d^k$ – for any i, $\binom{d}{i} \leq d^i$ and hence:

   $$N \leq d + d^2 + \ldots + d^k \leq \frac{d^{k+1}}{d-1} \leq 2d^k$$

   where the second inequality follows from the expression for the sum of a geometric series, and the third from the fact that $d \geq 2$. Since $|\mathcal{H}_k| \leq 2d^k$, so we get

   $$\text{err}(h) \leq \frac{1}{n} \ln \frac{|\mathcal{H}|}{\delta} = \frac{1}{n} \left(k \ln(2d) + \ln \frac{1}{\delta} \right).$$

2. *VC dimension.*

   (a) $VC(\mathcal{H}) = 2$. To prove this, we need two steps. We will first exhibit 2 points that $\mathcal{H}$ shatters, and second prove that there is no set of 3 points that $\mathcal{H}$ shatters.

   For the first part, consider $1, 2$. We can easily find intervals containing both and neither of these points. We can also find an interval $[0.5, 1.5]$ containing just 1, and $[1.5, 2.5]$ containing just 2. This covers all possibilities for labelling 1 and 2. Thus $\mathcal{H}$ shatters $\{1, 2\}$.

   Next, consider any set of 3 points. Without loss of generality, they are $a < b < c$. Observe that there is no interval that contains a and c but doesn't contain b. Therefore, it is impossible to have an interval achieve the labeled points $(a, 1)$, $(c, 1)$, and $(b, 0)$. This implies $\mathcal{H}$ does not shatter $\{a, b, c\}$. Since $\{a, b, c\}$ was arbitrary, we are done.

   (b) $VC(\mathcal{H}) = 4$. Consider the points $(1, 0), (0, 1), (-1, 0), (0, -1)$. These can be shattered by $\mathcal{H}$. Thus $VC(\mathcal{H}) \geq 4$. It now suffices to show that $VC(\mathcal{H}) < 5$. To do this, we must show that for any 5 points, $\mathcal{H}$ does not shatter them.

   Let S be a set of 5 points. Let a be the point with largest x-coordinate, b be the point with largest y-coordinate, c be the point with smallest x-coordinate, and d be the point with smallest y coordinate. Note that a, b, c, d are not necessarily distinct. Let e be any point in S that is not a, b, c, or d. Suppose for $\mathcal{H}$ shatters S. Then there must exist h such that $h(a) = h(b) = h(c) = h(d) = 1$ and $h(e) = 0$. However, this is impossible. A rectangle that contains a, b, c, d has an x-coordinate range and a y-coordinate range that encompass all of S, and thus include e, which gives us a contradiction.

3. It suffices to show that for any integer n that there exists a set of n points that $\mathcal{H}$ shatters. Fix n, and consider the set S of n evenly spaced points on the unit circle. We claim that $\mathcal{H}$ shatters this set.

   To prove this, consider any $T \subset S$. The key observation is that the convex hull of T includes all points in T and includes *no* points in $S \setminus T$. This is because we can connect the points of T in a convex polygon, and note that this polygon only intersects the unit circle at the points of T. Therefore the classifier in $\mathcal{H}$ corresponding to the convex hull of T (i.e. classifying all points in this region 1 and all other points 0) classifies T as 1 and $S \setminus T$ as 0. Since T was arbitrary, this shows $\mathcal{H}$ shatters S.

4. Let $C_1 = C_2$ both be the class of intervals described in problem 3(a). Then $\max(d_1, d_2) = 2$ since $d_1 = d_2 = 2$and $C_1 \cup C_2$ by our definition is the class of *double intervals* – $h_{[p,q],[r,s]}(x) = 1(p \leq x \leq q \text{ or } r \leq x \leq s)$.

   We will now show that this class has VC dimension more than two. Consider three points $a < b < c$ on the real line. We can see that $C = C_1 \cup C_2$ shatters $\{a, b, c\}$, since we can just take two intervals, one covering a and the other covering c to handle the case where a, b, c are labeled $1, 0, 1$. All other cases can be labeled with a single interval (meaning we pick the same interval from C_1 and C_2).

5. (a) Observe that for any u, $\sigma(u) - \sigma(-u) = \max(0, u) - \max(0, -u) = u$. Therefore, $a^\top x + b$ can be written as

   $$f(x) = 1 \cdot \sigma(a^\top x + b) + (-1) \cdot \sigma(-a^\top x - b)$$

   This corresponds to a single hidden layer neural network with two hidden nodes.

   (b) We claim that a one hidden layer neural network with $k = 2$ can achieve the function $|a^\top x + b|$; this is not a truncated linear function since it has two linear components – not a constant and a linear one. To show this, we first observe that for any u,

   $$\begin{aligned} \sigma(u) + \sigma(-u) = \max(0, u) + \max(0, -u) &= u, u \geq 0 \\ &= -u, u \leq 0 \end{aligned}$$

   Therefore, $\sigma(u) + \sigma(-u) = |u|$. Now we set $c_1 = c_2 = 1$, $a_1 = -a_2 = a$ and $b_1 = -b_2 = b$; then f becomes:

   $$f(x) = \sigma(a^\top x + b) + \sigma(-a^\top x - b) = |a^\top x + b|$$

6. The set of one-hidden-layer ReLU networks is defined by

   $$\mathcal{F} = \left\{ x \mapsto \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + e : k \in \mathbb{N}, a_i \in \mathbb{R}^d, b_i, c_i, e \in \mathbb{R} \right\}$$

   Take any $f \in \mathcal{F}$. f has the form

   $$f(x) = \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + e$$

   Then,

   $$g(x) = f(x) + b = \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + (e + b) \Rightarrow g \in \mathcal{F}$$

   $$h(x) = f(Dx) = \sum_{i=0}^{k} c_i \sigma(a_i^\top Dx + b_i) + e = \sum_{i=0}^{k} c_i \sigma((Da_i)^\top x + b_i) + e \Rightarrow h \in \mathcal{F}$$

   Therefore, $\mathcal{F}$ is closed under translation and scaling.