CSE 251A: Machine learning **Winter 2021**

Homework 5 — Generalization theory and Neural Networks

1. Let $\mathcal{X} = \{0,1\}^d$. The class $\mathcal{H}$ of *monotone disjunctions* consists of classifiers h that are given by a disjunction (logical OR) of some subset of the d features. For instance, the classifier

$$h(x) = x_1 \vee x_3 \vee x_8$$

assigns label 1 to points $x \in \mathcal{X}$ for which any of the features x_1, x_3, x_8 are set; and assigns label 0 otherwise. Suppose we obtain a training set of n points, drawn i.i.d. from an unknown underlying distribution, and we find a monotone disjunction $h \in \mathcal{H}$ that is correct on all n points. We would like to give a bound on the true error of h.

   (a) What is $|\mathcal{H}|$? Your answer should be a function of d.

   (b) Give a bound on the true error of h that holds with probability at least $1-\delta$ over the choice of training data.

   (c) What bound could you give if instead we looked at the smaller class $\mathcal{H}_k \subset \mathcal{H}$ of *k-sparse monotone disjunctions*: that is, monotone disjunctions consisting of at least 1 and at most k variables?

2. Determine the VC dimension of the following concept classes. Justify your answers.

   (a) *Intervals on the line.* $\mathcal{X} = \mathbb{R}$ and $\mathcal{H} = \{h_{a,b} : a, b \in \mathbb{R}, a < b\}$ where $h_{a,b}(x) = 1(a \le x \le b)$.

   (b) *Axis-aligned rectangles in the plane.* Each $h \in \mathcal{H}$ is given by an axis-aligned rectangle in $\mathbb{R}^2$, where points inside the rectangle are labeled 1, and points outside are labeled 0.

3. Let $\mathcal{X} = \mathbb{R}^2$, and let $\mathcal{H}$ be the hypothesis class of all convex sets on the plane. In other words, each $h_K \in \mathcal{H}$ corresponds to a convex set K, such that $h_K(x) = 1$ if x is inside or on the boundary of K and 0 otherwise. Show that $\mathcal{H}$ has infinite VC dimension.

4. Let C_1 be a concept class with VC dimension d_1 and C_2 be a concept class with VC dimension d_2. Let C be the concept class of unions of C_1 and C_2 – that is, each concept $c \in C$ corresponds to a concept $c_1 \in C_1$ and a $c_2 \in C_2$ such that:

$$c(x) = 1, \quad \text{iff } c_1(x) = 1 \text{or } c_2(x) = 1$$

Show an example of concept classes C, C_1 and C_2 such that the VC dimension of C is greater than $\max(d_1, d_2)$.

5. Suppose f is a single-hidden-layer neural network with d inputs, k hidden units and a single output unit. That is,

$$f(x) = \sum_{i=1}^{k} c_i \sigma(a_i^\top x + b_i) + e$$

where e, c_is and b_is are scalars, and each a_i is a $d \times 1$ vector. The non-linearity in the hidden units is a relu function – that is:

$$\begin{aligned} \sigma(u) &= u, \text{if } u \ge 0 \\ &= 0, \text{otherwise} \end{aligned}$$

(a) Show that any linear function $y = a^\top x + b$ of x can be represented by such an f.

(b) Suppose $k = 2$. A truncated linear function is a function of the form $g(x) = \max(a, b + c^\top x)$ – that is, a combination of a constant and a linear function. Give an example of a function that f can represent and that is not a truncated linear function.

6. Let $\mathcal{F}$ be the class of all one-hidden-layer relu networks on d inputs and a single output. Is $\mathcal{F}$ closed under translation – that is, for all $f \in \mathcal{F}$ and all scalars b, is $f(x) + b \in \mathcal{F}$? What about scaling – if $f(x) \in \mathcal{F}$, is $f(Dx) \in \mathcal{F}$ as well for any diagonal matrix D? Justify your answer.