Homework 3 TTIC 31250

Eddie Shim
eddieshim@uchicago.edu

05/06/2020

Exercise 1

I would generally like to explore some topic regarding the expressiveness of neural networks or the complexity-theoretic hardness of learning neural networks. Some ideas for a paper I may want to tackle include:

- Distribution Specific Hardness of Neural Networks; https://arxiv.org/pdf/1609.01037.pdf
- On the Complexity of Learning Neural Networks; https://papers.nips.cc/paper/7135-on-the-complexity-of-learning-neural-networks
- Benefits of Depth in Neural Networks; https://arxiv.org/abs/1602.04485
- The Power of Depth for Feedforward Neural Networks; https://arxiv.org/abs/1512.03965

Exercise 2

Consider the class H of axis-parallel rectangles in R^n. Specifically, a legal target function is specified by three intervals. Argue that $H[m] = O(m^6)$.

Consider the following set of points on a unit length rectangle that is aligned with the axis: (.5,.5,0), (0,.5,.5), (.5,.5,1), (1,.5,.5), (.5,0,.5), (.5,1,.5). These six points represent the middle point of each of the 6 sides of our rectangle. Notice that because each point lies on a unique side of our rectangle, we can include/exclude each point by shifting the walls of our rectangle either towards or away from the origin. Therefore, we know that the VC-dimension is at least 6. To prove that it cannot be greater than 6, notice that for our 7th point, if we place it in the interior of our bounding rectangle, then we cannot shatter all 7 points since we cannot achieve the hypothesis where the interior point is labeled negative but the rest of the 6 points are labeled positive. If there is no distinct interior point, then we must have that the 7th point lies on the same side as one of the other six points, thus we cannot achieve opposite predictions for the two points that lie on the same plane. Thus we know that the VC-dimension is 6.

Using Sauer's Lemma, we have:

$$H[m] \leq \sum_{i=0}^{VCdim(H)} \binom{m}{i} = O(m^{VCdim(H)})$$

$$H[m] \leq O(m^6)$$

Since the big O notation is an upper bound, we have our desired equality:

$$H[m] = O(m^6)$$

Problem 3

Show that TWO-LAYER$_{f,k}(H)$ has VC-dimension $O(kd\ \log kd)$. Note that we are only asking for an upper bound, not a lower bound.

Let the VC-dim of TWO-LAYER$_{f,k}(H)$ be represented by unknown variable m. Using Sauer's Lemma, we know that:

$$H[m] \leq O(m^d)$$

Thus, the number of ways to shatter m points using k functions can be represented by the inequality:

$$\binom{H[m]}{k} \leq m^{dk}$$

Thus, we have $2^m \leq m^{dk}$, which represents the final upper bound that we are trying to prove, that $m = O(kd\ \log kd)$

Problem 4

Prove that VC-dim$(H_n) \geq n + 1$ by presenting a set of $n + 1$ points in n- dimensional space such that one can partition that set with halfspaces in all possible ways.

Let our set of $n+1$ points be as such, take the point on the origin, and all other p_i points have coordinates all 0s and one 1, where the location of the 1 is where the ith coordinate on the ith point. For example, a 3-dimensional space, we have the set of points (0,0,0), (1,0,0), (0,1,0), (0,0,1). This outlines a unit-length cube similar to the one we worked with in exercise 2. In order to split this set of $n+1$ points into two subsets S_1, S_2, we look to find a set of hyperplanes H_n such that we can create any combination of points to create S_1, S_2 subsets. Since each coordinate is on the vertex of this cube, we can choose the following hyperplane when assuming the origin lies in S_1:

$$\sum_{i:p_i \in S_2} x_i = \frac{1}{2}$$

This hyperplane ensures that if you wish for your point p_i to be in S_2, then the hyperplane will intersect that dimension's axis at 0.5, thus separating p_i into S_2 from the origin and the rest of the points that are in S_1. Thus, since we know that we can produce a set of $n+1$ points that can be labeled in any way using H_n, it must be that VC-dim$(H_n) \geq n+1$.

Problem 5

Show that Radon's Theorem implies that the VC-dimension of halfspaces is at most $n+1$. Conclude that VC-dim$(H_n) = n+1$.

Radon's Theorem states that if we have S of $n+2$ points, then we can partition S into disjoint sets S_1, S_2 whose convex hulls intersect. Let point p be a point that exists in the intersection of both convex hulls. By definition of convex hull, we know that p can be written as $\sum_{x_i \in S'} \lambda_i x_i$, where each $\lambda_i \geq 0$, and $\sum_i \lambda_i = 1$. Since p is exists in both convex hulls, it can be written as either $\sum_{x_i \in S_1} \lambda_i x_i$ or $\sum_{x_i \in S_2} \lambda_i x_i$. We cannot draw any hyperplane that can separate S_1 on one side and S_2 on the other side, because that implies that the convex hull of S_1 is on one side and the convex hull of S_2 is on the other, but we know that p must be on both sides. The following proof by counter example demonstrates why. Assume $p \in S_1$ and that we can draw a hyper plane such that

$$a \cdot x_i \leq a_0; \forall x_i \in S_1$$

$$a \cdot x_i > a_0; \forall x_i \in S_2$$

However, we can show that p cannot be $\leq a_0$ because of the fact that it also exists in the convex hull of

S_2.

$$a \cdot p = \sum_{i:x_i \in S_2} \lambda_i a \cdot x_i > \left(\sum_{i:x_i \in S_2} \lambda_i \right) min_{i:x_i \in S_2}(a \cdot x_i) = min_{i:x_i \in S_2}(a \cdot x_i) > a_0$$

Therefore, because of this counterexample, we know that no set of $n+2$ points can be shattered. Since in problem 4 we proved that VC-dim $(H_n) \geq n+1$, we now have proved that VC-dim$(H_n) = n+1$.

Problem 6

Prove Radon's Theorem. As a suggested first step, prove that for any set of $n+2$ points $x_1, ..., x_{n+2}$ in $n-$dimensional space, there exist $\lambda_1, ..., \lambda_{n+2}$ not all zero such that $\sum_i \lambda_i x_i = 0$ and $\sum_i \lambda_i = 0$ (This is called an affine dependence).

First to prove affine dependence, consider a set of points S of $n+2$ points $x_1, ..., x_{n+2}$ and add another dimension with value 1 to each point, creating $(x_1, 1), ..., (x_{n+2}, 1)$. From the definition of convex hull, we know that we can find real valued $\lambda_1, ..., \lambda_{n+2}$, with each $\lambda_i \geq 0$, such that $\lambda_1(x_1, 1) + ... + \lambda_{n+2}(x_{n+2}, 1) = 0$. Thus this shows that we have a set of lambdas such that:

$$\sum_{i=1}^{n+2} \lambda_i x_i = 0$$

$$\sum_{i=1}^{n+2} \lambda_i = 0$$

Next to prove Radon's Theorem, let S be a set of $n+2$ points and let $\lambda_i, ..., \lambda_{n+2}$ be its affine dependence. We can split S into S_1, S_2, defining S_1 as the set of x_i points such that $\lambda_i > 0$¡ and S_2 as the set of points x_i such that $\lambda_i \leq 0$. Because of the properties of affine dependence, we know that

$$\sum_{i:x_i \in S_1} \lambda_i x_i = \sum_{i:x_i \in S_2} -\lambda_i x_i$$

Let $W_1 = \sum_{i:x_i \in S_1} \lambda_i$ and $W_2 = \sum_{i:x_i \in S_2} \lambda_i$. We know that $W_1 = -W_2$. Choose p as the following definition:

$$p = \sum_{i:x_i \in S1} \frac{\lambda_i}{W_1} x_i = \sum_{i:x_i \in S_2} \frac{\lambda_i}{W_2} x_i$$

Since our chosen p exists in both convex hulls of S_1 and S_2, we have proven that the convex hulls intersect. Thus, we have proven that a set of $n+2$ points in n dimensions can be partitioned into two disjoint subsets S_1, S_2 whose convex hulls intersect.