Mehryar Mohri
Foundations of Machine Learning 2014
Courant Institute of Mathematical Sciences
Homework assignment 2
October 3, 2014
Due: October 17, 2014

A. VC-dimension of axis-aligned squares or triangles

1. What is the VC-dimension of axis-aligned squares in the plane?

   Solution: It is not hard to see that the set of 3 points with coordinates $(1, 0)$, $(0, 1)$, and $(-1, 0)$ can shattered by axis-aligned squares: e.g., to label positively two of these points, use a square defined by the axes and with those to points as corners. Thus, the VC-dimension is at least 3. No set of 4 points can be fully shattered. To see this, let P_T be the highest point, P_B the lowest, P_L the leftmost, and P_R the rightmost, assuming for now that these can be defined in a unique way (no tie) – the cases where there are ties can be treated in a simpler fashion. Assume without loss of generality that the difference d_{BT} of y-coordinates between P_T and P_B is greater than the difference d_{LR} of x-coordinates between P_L and P_R. Then, P_T and P_B cannot be labeled positively while P_L and P_R are labeled negatively. Thus, the VC-dimension of axis-aligned squares in the plane is 3.

2. Consider right triangles in the plane with the sides adjacent to the right angle both parallel to the axes and with the right angle in the lower left corner. What is the VC-dimension of this family?

   Solution: It is not hard to see that the set of 3 points with coordinates $(0, 0)$, $(-1, -1)$, $(-2, -2)$, and $(-3, -3)$ can shattered by such triangles. To see that no five points can be shattered, the same example or argument as for axis-aligned rectangles can be used: labeling all points positively except from the one within the interior of the convex hull is not possible (for the degnerate cases where no point is in the interior of the convex hull is simpler, this is even easier to see). Thus, the VC-dimension of this family of triangles is 4.

B. Growth function bound

1. Consider the family H of threshold functions over $\mathbb{R}^N$ defined by $\{\mathbf{x} = (x_1, \ldots, x_N) \mapsto \operatorname{sgn}(x_i - \theta) \colon i \in [1, N], \theta \in \mathbb{R}\}$, where $\operatorname{sgn}(z) = +1$ if $z \geq 0$, $\operatorname{sgn}(z) = -1$ otherwise. Give an explicit upper bound on the growth function $\Pi_H(m)$ of H that is in $O(mN)$.

   Solution: For each feature, there at most $m + 1$ ways of selecting the threshold (between any two feature values or beyond or below all values). Thus, the total number of thresholds functions for a sample of size m is at most $(m + 1)N$. Thus, the growth function is upper bounded by $(m + 1)N$.

2. In class, we gave a bound on the Rademacher complexity of a family G in terms of the growth function (Lecture 3, slide 18). Show that a finer upper bound on the Rademacher complexity can be given in terms of $\mathrm{E}_S[\Pi(G, S)]$, where $\Pi(G, S)$ is the number of ways to label the points in sample S.

   Solution: Following the proof given in class and using Jensen's inequality (at the last step), we can write:

$$
\begin{aligned}
\widehat{\mathfrak{R}}_m(G) &= \mathop{\mathrm{E}}_{S,\sigma}\left[\sup_{g \in G} \frac{1}{m} \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_m \end{bmatrix} \cdot \begin{bmatrix} g(z_1) \\ \vdots \\ g(z_m) \end{bmatrix}\right] \\
&\leq \mathop{\mathrm{E}}_{S}\left[\frac{\sqrt{m}\sqrt{2 \log |\{(g(z_1), \ldots, g(z_m)) \colon g \in G\}|}}{m}\right] \qquad \text{(Massart's Lemma)} \\
&= \mathop{\mathrm{E}}_{S}\left[\frac{\sqrt{m}\sqrt{2 \log \Pi(G, S)}}{m}\right] \\
&\leq \frac{\sqrt{m}\sqrt{2 \log \mathrm{E}_S[\Pi(G, S)]}}{m} = \sqrt{\frac{2 \log \mathrm{E}_S[\Pi(G, S)]}{m}}.
\end{aligned}
$$

C. VC-dimension of neural networks

Let C be a concept class over $\mathbb{R}^r$ with VC-dimension d. A C-neural network with one intermediate layer is a concept defined over $\mathbb{R}^n$ that can be represented by a directed acyclic graph such as that of Figure 1, in which the input nodes are those at the bottom and in which each other node is labeled with a concept $c \in C$.

The output of the neural network for a given input vector $(x_1, \ldots, x_n)$ is obtained as follows. First, each of the n input nodes is labeled with the corresponding value $x_i \in \mathbb{R}$. Next, the value at a node u in the higher layer and labeled with c is obtained by applying c to the values of the input nodes admitting an edge ending in u. Note that since c takes values in $\{0,1\}$, the value at u is in $\{0,1\}$. The value at the top or output node is obtained similarly by applying the corresponding concept to the values of the nodes admitting an edge to the output node.

1. Let H denote the set of all neural networks defined as above with $k \geq 2$ internal nodes. Show that the growth function $\Pi_H(m)$ can be upper bounded in terms of the product of the growth functions of the hypothesis sets defined at each intermediate layer.

   Solution: Let $\Pi_u(m)$ denote the growth function at a node u in the intermediate layer. For a fixed set of values at the intermediate layer, using the concept class C the output node can generate at most $\Pi_C(m)$ distinct labelings. There are $\prod_u \Pi_u(m)$ possible sets of values at the intermediate layer since, by definition, for a sample of size m, at most $\Pi_u(m)$ distinct values are possible at each u. Thus, at most $\Pi_C(m) \times \prod_u \Pi_u(m)$ labelings can be generated by the neural network and $\Pi_H(m) \leq \Pi_C(m) \prod_u \Pi_u(m)$.

2. Use that to upper bound the VC-dimension of the C-neural networks (*hint*: you can use the implication $m = 2x \log_2(xy) \Rightarrow m > x \log_2(ym)$ valid for $m \geq 1$, and $x, y > 0$ with $xy > 4$).

   Solution: For any intermediate node u, $\Pi_u(m) = \Pi_C(m)$. Thus, $\Pi_H(m) \leq \Pi_C(m)^k$. By Sauer's lemma, $\Pi_C(m)^k \leq \left(\frac{em}{d}\right)^d$, thus $\Pi_H(m) \leq \left(\frac{em}{d}\right)^{dk}$. Let $m = 2kd \log_2(ek)$. In view of the inequality given by the hint and $ek > 4$, this implies $m > dk \log_2\left(\frac{em}{d}\right)$, that is $2^m > \left(\frac{em}{d}\right)^{dk}$. Thus, the VC-dimension of H is less than

   $$2kd \log_2(ek).$$

3. Let C be the family of concept classes defined by threshold functions $C = \{\mathrm{sgn}(\sum_{j=1}^{r} w_j x_j) \colon \mathbf{w} \in \mathbb{R}^r\}$. Give an upper bound on the VC-dimension of H in terms of k and r.

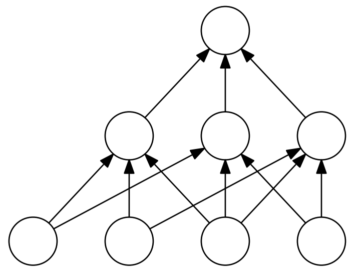

Figure 1: A neural network with one intermediate layer.

Solution: For threshold functions, the VC-dimension of C is r, thus, the VC-dimension of H is upper bounded by

$$2kr \log_2(ek).$$