Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 2 – Solution
February 21, 2006

Problem 1: VC dimension [75 points]

(1) [25 points]

(a) [5 points] It suffices to show the existence of a set of $n+1$ points in $\mathbb{R}^n$ that can be shattered by halfspaces. Let x_0 be the origin and define x_i as the point whose ith coordinate is 1 and all others 0. Let $y_0, y_1, \ldots, y_n \in \{-1, +1\}$ be an arbitrary set of labels for $x_0, \ldots, x_n$. Let w be the vector whose ith coordinate is y_i. Then, the classifier defined by the hyperplane of equation $w \cdot x + y_0/2 = 0$ shatters $x_0, \ldots, x_n$ since:

$$\begin{aligned} &\text{sign}(w \cdot x_0 + y_0/2) = \text{sign}(y_0 + y_0/2) = y_0, \text{and} \\ &\forall\, i \geq 1,\ \text{sign}(w \cdot x_i + y_0/2) = \text{sign}(y_i + y_0/2) = y_i. \end{aligned} \tag{1}$$

(b) [10 points] It suffices to show that no set of $n+2$ points can be shattered by halfspaces. Let X be a set of $n+2$ points. By Radon's theorem, it can be split into two sets X_1 and X_2 such that their convex hulls intersect.

Observe that when two sets of points X_1 and X_2 are separated by a hyperplane, their convex hulls are also separated by that hyperplane. Thus, X_1 and X_2 cannot be separated by a hyperplane and X is not shattered.

(c) [10 points] Let $I_1 = \{i \in [1, n+2] : x_i \in X_1\}$ and $I_2 = \{i \in [1, n+2] : x_i \in X_2\}$. x is in the convex hull of X_1 and X_2 iff there exist $(\alpha_i)_{i \in I_1}$ and $(\alpha_i)_{i \in I_2}$ such that

$$\begin{aligned} x &= \sum_{i \in I_1} \alpha_i x_i \text{ with } \sum_{i \in I_1} \alpha_i = 1, \text{ and} \\ x &= \sum_{i \in I_2} \alpha_i x_i \text{ with } \sum_{i \in I_2} \alpha_i = 1. \end{aligned} \tag{2}$$

This leads to the following system of $n+1$ equations in $n+2$ unknown α_i:

$$\begin{cases} \sum_{i \in I_1} \alpha_i x_i - \sum_{i \in I_2} \alpha_i x_i = 0 \\ \sum_{i \in I_1} \alpha_i - \sum_{i \in I_2} \alpha_i = 0, \end{cases} \tag{3}$$

which has a non-trivial solution. This proves Radon's theorem.

(2) [30 points] Let $m \geq 0$. Note the general fact that for any concept class $C = \{c_1 \cap c_2 : c_1 \in C_1, c_2 \in C_2\}$,

$$\Pi_C(m) \leq \Pi_{C_1}(m)\, \Pi_{C_2}(m). \tag{4}$$

Indeed, fix a set X of m points. Let $Y_1, \ldots, Y_k$ be the traces of C_1 on X. By definition of $\Pi_{C_1}(X)$, $k \leq \Pi_{C_1}(X) \leq \Pi_{C_1}(m)$. By definition of $\Pi_{C_2}(Y_i)$, The traces of C_2 on a subset Y_i are at most $\Pi_{C_2}(Y_i) \leq \Pi_{C_2}(m)$. Thus, the traces of C on X are at most

$$k\Pi_{C_2}(Y_i) \leq \Pi_{C_1}(m)\, \Pi_{C_2}(m). \tag{5}$$

For the particular case of C_k, using Sauer's lemma, this implies that

$$\Pi_{C_k}(m) \leq (\Pi_{C_1}(m))^k \leq \left(\frac{em}{n+1}\right)^{k(n+1)}. \tag{6}$$

If $(em/(n+1))^{k(n+1)} < 2^m$, then the VC dimension of C_k is less than m. If the VC dimension of C_k is m, then $\Pi_{C_k}(m) = 2^m \leq (em/(n+1))^{k(n+1)}$. These inequalities give an upper bound and a lower bound on $\text{VCdim}(C_k)$. As an example, using the identity: $\forall x \in \mathbb{N} - \{3\}, \log_2(x) \leq x/2$, one can verify that:

$$\text{VCdim}(C_k) \leq 2(n+1)k\log(3k). \tag{7}$$

(3) [20 points]

(a) [5 points] When $C = A \cup B$, $\Pi_C(X) \leq \Pi_A(X) + \Pi_B(X)$ for any set X since dichotomies in $\Pi_C(X)$ can be generated by A or by B. Thus, for all m, $\Pi_C(m) \leq \Pi_A(m) + \Pi_B(m)$.

(b) [15 points] For $m \geq d_A + d_B + 2$, by Sauer's lemma,

$$\begin{aligned}
\Pi_C(m) &\leq \sum_{i=0}^{d_A}\binom{m}{i} + \sum_{i=0}^{d_B}\binom{m}{i} = \sum_{i=0}^{d_A}\binom{m}{i} + \sum_{i=0}^{d_B}\binom{m-i}{i} \\
&= \sum_{i=0}^{d_A}\binom{m}{i} + \sum_{i=m-d_B}^{d_B}\binom{m}{i} \qquad (8) \\
&\leq \sum_{i=0}^{d_A}\binom{m}{i} + \sum_{i=d_A+2}^{d_B}\binom{m}{i} \qquad (9) \\
&< \sum_{i=0}^{m}\binom{m}{i} = 2^m. \qquad (10)
\end{aligned}$$

Thus, the VC dimension of C is strictly less than $d_A + d_B + 2$:

$$\text{VCdim}(C) \leq d_A + d_B + 1. \tag{11}$$

Is this bound tight (can you show that for any d_A and d_B, there exist sets A and B such that equality holds)?

Problem 2: Sample complexity [25 points]

(a) [15 points] For $i = 0, \ldots, n$, let $x_i \in \{0,1\}^n$ be defined by $x_i = (\underbrace{1, \ldots, 1}_{i \text{ 1's}}, 0, \ldots, 0)$. Then, $\{x_0, \ldots, x_n\}$ can be shattered by C. Indeed, let $y_0, \ldots, y_n \in 0, 1$ be an arbitrary labeling of these points. Then, the function h defined by:

$$h(x) = y_i \tag{12}$$

for all x with i 1's is symmetric and $h(x_i) = y_i$. Thus, $\text{VCdim}(C) \geq n + 1$. Conversely, a set of $n + 2$ points cannot be shattered by C since at least two points would then have the same number of 1's and will not be distinguishable by C. Thus,

$$\text{VCdim}(C) = n + 1. \tag{13}$$

(b) [5 points] Thus, in view of the theorems presented in class, a lower bound on the number of training examples needed to learn symmetric functions with accuracy $1 - \epsilon$ and confidence $1 - \delta$ is

$$\Omega(\frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{n}{\epsilon}), \tag{14}$$

and an upper bound is:

$$O(\frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{n}{\epsilon} \log \frac{1}{\epsilon}), \tag{15}$$

which is only within a factor $\frac{1}{\epsilon}$ of the lower bound.

(c) [5 points] This is trivial. For a training data $(z_0, t_0), \ldots, (z_m, t_m) \in \{0,1\}^n \times \{0,1\}$ define h as the symmetric function such that $h(z_i) = t_i$ for all $i = 0, \ldots, m$.

Can you show that in view of the bounds given in (b), this algorithm is optimal?