Homework 8 - Solutions

Problem 1

Let $X_n = \{1, \ldots, n\}$. Let $C_k$ be the set of all subsets of $X_n$ of size $k$.

Now, we split into two cases:

1. $n \geq 2k$: Let $S = \{1, \ldots, k\}$. So $|S| = k$, and $|X_n| = n \geq 2k$. Thus, $\forall S_i \subseteq S$ (there are $2^k$ subsets of $S$), we can construct $T_i$ s.t. $T_i = S_i \cup \left\{x \mid k + 1 \leq x \leq k + (k - |S_i|)\right\}$. Note that since $S_i$ and $\left\{x \mid k + 1 \leq x \leq k + |S_i|\right\}$ are disjoint and their sizes add up to $k$, then $T_i \in C_k$. Thus, $T_i \cap S_i = S_i$, $\forall S_i \subseteq S$, so $C_k$ shatters $S$, so the VC-dimension is at least $k$.

Clearly, it is impossible for an element in $C_k$ to shatter any subset of $X_n$ of size $k + 1$ since all the elements of $C_k$ have size $k < k + 1$, so no set in $C_k$ can intersect entirely with all $k + 1$ elements. Thus, the VC-dimension is $k$ in this case.

2. $n < 2k$: Let $S = \{1, \ldots, n - k\}$. So $|S| = n - k$, and $|X_n| = n < 2k$. Thus, $\forall S_i \subseteq S$ (there are $2^{n-k}$ subsets of $S$), we can construct $T_i$ s.t. $T_i = S_i \cup \left\{x \mid n - k + 1 \leq x \leq n - |S_i|\right\}$. Note that as above, the two constituent subsets of $T_i$ are disjoint, so $|T_i| = |S_i| + \left\{x \mid n - k + 1 \leq x \leq n - |S_i|\right\}$.

Thus, $T_i \cap S_i = S_i$, $\forall S_i \subseteq S$, so $C_k$ shatters $S$, thus the VC-dimension in this case is at least $n - k$.

However, it is impossible for an element in $C_k$ to shatter any subset of $X_n$ of size $n - k + 1$, call an arbitrary such subset $S_{n-k+1}$, because $\emptyset \subseteq S_{n-k+1}$. In order for an element $C \in C_k$ to satisfy $C \cap S_{n-k+1} = \emptyset$, then $C$ and $S_{n-k+1}$ must be disjoint, however this is impossible as this would imply that $|X_n| \geq n + 1$. Thus, the VC-dimension is $n - k$ in this case.

Therefore, the VC-dimension for $(X_n, C_k)$ is $\min\{k, n - k\}$.

Problem 2

a) Let $\mathcal{R}$ be the set of axis-aligned rectangles in $\mathbb{R}^2$.

i. VC dimension is at least 4. Let $S = \{(-1,0), (0,1), (1,0), (0,-1)\}$. $S$ is clearly shattered by $\mathcal{R}$ as you can draw a rectangle around any subset of the points in $S$. This can be seen by picture which is elided.

ii. VC dimension is at most 4. Consider an arbitrary set of 5 points, $S$. Let $x_<$ be a point with minimal $x$ coordinate, and let $x_>$ be a point with maximal $x$ coordinate (breaking ties arbitrarily). Define $y_<$ and $y_>$ similarly except for $y$ coordinates. Note that $|T| = |\{x_<, x_>, y_>, y_<\}| \leq 4$, so $|S \setminus T| \geq 1$. By definition of these points, any axis aligned rectangle that contains the points in $T$ must necessarily contain all points in $S$ as any points in $S$ must have coordinates within these extrema. Therefore, it is impossible to classify the points in $T$ positively and the points in $S \setminus T$ negatively. We have shown that $S \setminus T \neq \emptyset$, therefore $S$ is not shattered by $\mathcal{R}$. 

b) Consider the following algorithm:

1) Take \( n \) samples from the distribution \( D \) and label them using the given Oracle
2) Let \( S \) be the set of points that are labeled positive
3) If \( S \) is empty, return \( \emptyset \). Otherwise, output the tightest axis-aligned rectangle, \( R' \), containing the points in \( S \). This is the rectangle with edges corresponding to the points \( \{x_>, x_>, y_<, y_<\} \) as defined above.

We will show that this algorithm satisfies the PAC learning conditions for \( n \geq \frac{4}{\epsilon} \log \frac{4}{\delta} \) samples.

Fix \( 0 < \epsilon, \delta \leq 1 \). Define \( R(x) \) to be the classification of a point \( x \) by a rectangle \( R \). Let \( R \in \mathcal{R} \) be the true concept and let \( R' \in \mathcal{R} \) be the rectangle that is output by the above algorithm. Note that since we choose the smallest rectangle correctly classifying the positive samples, \( R' \subseteq R \), as any positive sample must be contained in \( R \). We will now show that with probability at least \( 1 - \delta \), the algorithm above returns a rectangle \( R' \) whose error is at most \( \epsilon \). Define \( \text{err}_P(R') = P_D(R(x) \neq R'(x)) \).

First, consider the case where \( P_D(x \in R) \leq \epsilon \). We know that \( \text{err}_P(R') \leq P_D(x \in R) \leq \epsilon \) with probability \( 1 \), so \( R \) is PAC learnable.

Now, consider the case where \( P_D(x \in R) > \epsilon \). Note that since \( R \) and \( R' \) are axis-aligned rectangles and \( R' \subseteq R \), the only region where \( R' \) can differ from \( R \) is the area \( R \setminus R' \), which is the area between \( R' \) and \( R \).

Let \( X_< \) be a rectangular region, contained in \( R \), starting at the left edge of \( R \) such that \( P_D(X_<) = \frac{\epsilon}{4} \) and \( X_> \) be the analogous region for the right edge of \( R \). Let \( Y_< \) be the same for the bottom edge of \( R \) and \( Y_> \) be the same for the top edge of \( R \). If the distribution \( D \) is assumed to be smooth enough to not contain any points with point mass, these regions must exist. For simplicity, we assume this is the case.

If at least one sample fell in each one of \( X_<, X_>, Y_<, Y_> \), then \( R' \) would have error at most \( (\text{by the union bound}) P_D(X_< \cup X_> \cup Y_< \cup Y_>)) \leq P_D(X_<) + P_D(X_> + P_D(Y_<) + P_D(Y_>) = 4 \cdot \frac{\epsilon}{4} = \epsilon \).

Thus, for \( R' \) to have error more than \( \epsilon \), it must be the case the at least one of these four regions did not see any samples. Let \( E_{x,<} \) be the event that no samples fell in \( X_< \) and similarly define the events \( E_{x,>}, E_{y,<}, E_{y,>} \). Thus, we have that, by the union bound,

\[
P(\text{err}_P(R') > \epsilon) \leq P(E_{x,<} \cup E_{x,>} \cup E_{y,<} \cup E_{y,>}) \leq P(E_{x,<}) + P(E_{x,>}) + P(E_{y,<}) + P(E_{y,>})
\]

Note that \( P(E_{x,<}) = P_D(\bigcap_{j=1}^{n} \{s_i \notin X_<\}) = (1 - \frac{\epsilon}{4})^n \leq e^{-n \epsilon/4} \) and similar bounds hold for the other three events. Thus, \( P(\text{err}_P(R') > \epsilon) \leq 4e^{-n \epsilon/4} \). So for \( n \geq \frac{1}{\epsilon} \log \frac{4}{\delta} \) samples, we get that \( P(\text{err}_P(R') > \epsilon) \leq \delta \) and therefore the PAC-learning guarantee holds.