1. Do two of the three following VC dimension questions. Identify the question at the start of your answer. Provide both a shattered set and a proof that no larger set is shattered.

   (a) Let the domain be $\mathbb{R}^2$ (the plane) and $\mathcal{H}$ be the set of triangles. In other words, there is a correspondence between each $h \in \mathcal{H}$ and each triple of points $a, b, c$ (the vertices of the triangle) so that $x \in h$ if and only if there exists $w_1, w_2, w_3$ such that each $w_i \geq 0$, $w_1 + w_2 + w_3 = 1$, and $x = w_1a + w_2b + w_3c$ (this formal definition may not be useful in solving the problem).

   (b) Let the domain be $\mathbb{R}^d$ ($d$-dimensional euclidean space) and let $\mathcal{H}$ be the set of homogeneous half-spaces. Thus there is a correspondence between hypotheses $h \in \mathcal{H}$ and vectors $w \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$ is in an $h$ iff $x \cdot w \geq 0$ (for the corresponding $w$). You may use the fact that in any set $S$ of $d + 1$ points in $\mathbb{R}^d$, there is at least one point $x \in S$ that can be expressed as a linear combination of the other points in $S$.

   (c) Let the domain be $\{0, 1\}^d$ (represent $d$ boolean features that can be either true or false) and let the hypothesis class $\mathcal{H}$ be the set of conjunctions of these $d$ boolean features. Thus each $h \in \mathcal{H}$ is isomorphic to a subset of the $d$ features, and an $x \in h$ if and only if $x$ has the value 1 for each feature in the subset.

2. Experiment with how many examples are needed to closely approximate intervals in the following three situations. Compare your results with the PAC learning bounds. In each case, consider the domain $[0, 1]$ and a target interval like $[0.4, 0.6]$. You may approximate the continuous interval with a large number of bins (say 1000) and assume that the distribution is uniform over the bins. Be sure to describe your experimental methodology and give an indication of how much larger the error of a most accurate hypothesis interval on the data is than than the error of the best hypothesis on the distribution (including noise) as a function of the number of training examples. Consider the confidence $\delta$ fixed to something like 0.9 and emphasize the accuracy ($\epsilon$) values you can get with different sample sizes.

   (a) The noise free case – examples are positive if they fall in the target interval and negative otherwise.

   (b) With iid label noise. Each example is labeled as above, but then there is a chance (perhaps 20%) that the label is flipped.

   (c) With attribute noise. After labeling the point, add a random number chosen uniformly from $[-0.2, +0.2]$ to the point so it can shift a number of bins left or right.

This problem is a little open ended, and it would be interesting to find other simple noise settings (or perhaps a target concept $c \notin \mathcal{H}$) such that many examples are needed
for intervals with the best performance on the training data to closely approach the predictive performance of the best interval on new data.

3. Problem 7.2 in the text: Consider the class $C$ of concepts of the form $(a \leq x \leq b) \land (c \leq y \leq d)$, where $a, b, c, d$ are integers in the interval $(0, 99)$ inclusive. Note that each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the $x,y$ plane. Hint: Given a region in the plane bounded by the points $(0,0)$ and $(n-1,n-1)$, the number of distinct rectangles with integer-valued boundaries within this region is $\left(\frac{n(n+1)}{2}\right)^2$.

(a) Give an upper bound on the number of randomly drawn training examples sufficient to assure, that for any target concept $c$ in $C$, any consistent learner using $H = C$ will, with probability $95\%$, output a hypothesis with error at most 0.15.

(b) Repeat the above assuming that the rectangle boundaries take on real values instead of integers.