Donovan Guelde
CSCI 5622
Learning HW

1. To determine the number of training samples required to find a hypothesis that, with probability 95%, has an error of at most 0.15, we use the formula $m \geq \frac{1}{\epsilon}(n\ ln(|H|)+ln(\frac{1}{\delta}))$, where:

$\epsilon = .15$, $\delta = .05$ and $|H|$ is the number of hypotheses in our hypothesis class.

$|H|$ can be determined from the geometry of our hypothesis space, that of all possible triangles with vertexes (i,j) where i,j$\in$ integers in the range [0,99].

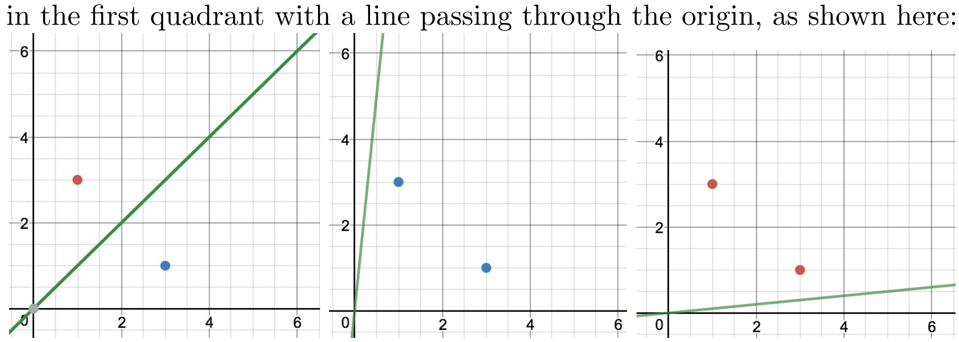
For each vertex of the triangle, there are (100*100) possible values for (i,j), and there are three vertexes, so the number of hypotheses in our hypothesis class is $(100^2)^3$, or 1×10^{12}.

Now we have:

$m \geq \frac{1}{0.15}(ln(1 \times 10^{12}) + ln(\frac{1}{0.05})) = 572.592$

To achieve the desired probability and accuracy, we need at least 573 training samples.

2. Considering the hypothesis class H of linear hyperplanes in 2D that pass through the origin, the lower bound on the VC dimension is 2. It is possible to shatter a set of two points in the first quadrant with a line passing through the origin, as shown here:



It appears that the upper bound of this hypothesis class is also 2, but we must prove it. Do do so, we must show than no set of points $p_1, p_2, p_3 = (x_1, y_1), (x_2, y_2), (x_3, y_3)$ can be shattered by a line passing through the origin.

First, the points p_1, p_2, and p_3 are placed anywhere in the first quadrant. Our classifier, the line passing through the origin, begins at an angle θ from the origin. Define our classification scheme as points above the line are labeled as (+), and points below the line are labeled (-). At $\theta = 0$, the line is on the x-axis, and will classify all three points the same (+,+,+). Now we increase angle θ until the line crosses one and only one point. Now, one point is separated from the other two (representing a (+,+,-) classification). Continuing to increase the angle θ, the line will cross over the second point, p_2, resulting in a labelling of (+,-,-). Again increasing θ until it crosses the point p_3, we have the classification (-,-,-). Ordered by increasing angle θ required to achieve the specific classification, we have $(+,+,+) \leq (+,+,-) \leq (+,-,-) \leq (-,-,-)$. We use $\leq$ in the previous inequality to account for the possibility that two or more points may be colinear relative to our classifying line.

These are the only possible outcomes of our classification hypothesis class, since as θ increases from 0 to $\frac{\pi}{2}$ radians, it is impossible to classify a three-point system via our hypothesis class as (+,-,+) or (-,+,-). Therefore, the upper bound on the VC dimension is 2, the lower bound is 2, as shown previously, VC dimention is 2.

3. To calculate the weight w that allows our classification hypothesis $sign(sin(w * \frac{1}{2^{-k}}))$ to shatter any given set of points on the x-axis, it is helpful to recall some fundamental properties of the sin function:

$sin(\pi x) = -sin(x)$
$sin(2\pi x) = sin(x)$ (holds true for any even multiple of π)
$sin(x)$ is positive for all values (0,π), and negative for all values of i (π,2π)

To shatter any given set of points $x_1, x_2, ..., x_n$ with labels $y_1, y_2, ..., y_n$, we can use the following summation to determine weight w:

$w = \pi(1 + \sum_i^n (1 - y_i)(2^{x_i}))$
and we use this w in our classifier: $y_i = sign(sin(\frac{w}{2^k}))$

When the above function is evaluated, $\frac{w}{2^{-k}}$ will be simplified in one of two ways:

Case 1: Where y_i = True:
Since all of the terms involved in the summation for determining w are derived from the $y_i = False$ training examples, the term in question will simplify to:
$\frac{\pi}{2^k}$ + (some multiple of 2 * π), giving the classifier as:
$sign(sin(\frac{1}{2^z}\pi))$ (where z is come integer $\geq$ 1), which is always positive.

Case 2: Where $y_i = False$
If $y_i = False$, then the term $(\frac{w}{2^k})$ simplifies to:
$\frac{\pi}{2^k}$ + π + (some multiple of 2* π), which is always in the range$(\pi, 2\pi)$.
This causes the classifier to return (-)