CMSC726 : Machine Learning *Ravi Bhadeshiya*

Learnability Analysis

1. Argue about an ordering of hypothesis classes in terms of complexity: hyperplanes through the origin, arbitrary hyperplanes, and axis-aligned rectangles
   Ans: The experiment was done with 20 random points and the Rademacher estimate are as follow:

```
Rademacher correlation of plane classifier 0.320000
0.03125
Rademacher correlation of extra plane classifier 0.460000
0.1875
Rademacher correlation of rectangle classifier 0.640000
354.078125

Process finished with exit code 0
```

   From this result, Order of complexity can be specify by Rademacher estimate which is as follow in acceding order :
   Hyperplane through origin < Hyperplane through arbitrary plane < Axis-aligned rectangles

   This experimental result is very much related to VC-dimensions of each hypothesis. For Hyperplane passed through origin, VC-dimensions is 2 means this can't shatter the 3 points and it just need to tune slope of line which is only 1. Intuitively it becomes less complex. For the hyperplane passed through arbitrary points, VC-dimension is 3 and It need to tune slope and intercept of the lines which makes it more complex than the origin-centered hyperplanes. For Axis-aligned rectangles, VC-dimension is 4 and it cannot shatter 5 points.Since it need to tune four parameter - $X_{min}, Y_{min}, X_{max}, Y_{max}$ - makes it most complex than other. All in other word, one could expect a learning machine with high parameter will have high VC-dimension, whereas machine with few parameter will have low VC-dimension.

   In conclusion, The Rademacher and VC-dimension are different measurements of complexity learning machine.

2. Prove that your frequency correctly classifies any training set (up to floating point precision on the computer).
   Ans:
   The reference for this solution is taken from: `http://mlweb.loria.fr/book/en/VCdiminfinite.html`,

   The set of classifier can be selected as equation(1) for a labeled dataset of the form $\{(2^{-i}, y_i)\}_{i=1}^{n}$ and for any similar dataset.

$$h(x_j) = sign\left(\sin\left(\omega x_j\right)\right) \quad ; \omega = \pi\left(1 + \sum_{i=1} \frac{1-y_i}{2} 2^i\right) \tag{1}$$

   For a given classification it observe that, for any point $x_j = 2^{-j}$ in the considered data set such that $y_j = -1$, the term 2^j appears in the sum. This leads to..

$$\begin{aligned} h(x_j) &= sign\left(\sin\left(\pi 2^{-j}\left(1 + \sum_{i=1} \frac{1-y_i}{2} 2^i\right)\right)\right) \\ &= sign\left(\sin\left(\pi 2^{-j} + \frac{1-y_i}{2}\pi + \sum_{\{i\ :\ i>j\}} \frac{1-y_i}{2}\pi 2^{i-j} + \sum_{\{i\ \ i<j\}} \frac{1-y_i}{2}\pi 2^{i-j}\right)\right) \end{aligned}$$

$$\text{Where, } \omega x_j = \pi \left(2^{-j} + 1 + \sum_{\{i\,:\,y_i=-1,\ i>j\}} 2^{i-j} + \sum_{\{i\,:\,y_i=-1,\ i<j\}} 2^{i-j} \right)$$

For $\forall i > j$, the terms 2^{i-j} are positive powers of 2 and thus are even numbers that can be written as $2k_i$ for some $k_i \in \mathbb{N}$. Regarding the other remaining sum where $\forall i < j$, can be defined as geometric series which gives ωx_j as follow:

$$\omega x_j = \pi\,(1+\epsilon) + 2k\pi$$

Given that $2^{-j} \leq 0.5$,

$$\pi < \pi(1+\epsilon) < 2\pi \quad \Rightarrow \quad \sin(\omega x_j) < 0$$

Hence, the classifier correctly predicts all negative labels $y_j = -1 = \text{sign}(\sin(\omega x_j)) = h(x_j)$ The same steps can be reproduce with positive labels $y_j = +1$ with the difference that the term 2^j does not appear in the sum defining ω. This leads to

$$\begin{aligned} \omega x_j &= \pi \left(2^{-j} + \sum_{\{i\,:\,y_i=1,\ i>j\}} 2^{i-j} + \sum_{\{i\,:\,y_i=1,\ i<j\}} 2^{i-j} \right) \\ &= \pi\epsilon + 2k\pi \end{aligned}$$

$$\text{With, } 0 < \pi\epsilon < \pi \quad \Rightarrow \quad \sin(\omega x_j) > 0$$

Thus, all positively labeled points are also correctly classified by h using the particular choice of ω.

3. Suppose we are classifying real numbers, not integers. The classifier returns positive (1) if the point is greater than the sin function and negative (0) otherwise. Give an example of four points that cannot be shattered by this classifier. How does this relate to the VC dimension?

Ans:

For some real number k with equally spaced points such that $x = \{k, 2k, 3k, 4k\}$ with labels, $y = \{+, +, -, +\}$, there is some ω it can't be shattered.Thus,

$$sin(\omega k) \geq 0, \quad sin(2\omega k) < 0, \quad sin(3\omega k) < 0, \quad sin(4\omega k) < 0$$

Now,

$$\begin{aligned} Sin(4k\omega) &= 2 \cdot sin(2k\omega) \cdot cos(2k\omega) \geq 0 \\ &\Rightarrow 2 \cdot sin(2k\omega) \cdot (1 - 2sin^2(k\omega)) \geq 0 \quad ;\text{Where } sin(2k\omega) \geq 0 \end{aligned}$$

$$\text{Thus, } (1 - 2sin^2(k\omega)) \geq 0 \Rightarrow sin^2(k\omega) < \frac{1}{2}$$

While consider this,

$$\begin{aligned} Sin(3k\omega) &= 3\sin(k\omega) - 4\sin^3(k\omega) < 0 \\ &\Rightarrow 2 \cdot \sin(k\omega)(3 - 4\sin^2(k\omega) < 0 \quad ;\text{Where } sin(k\omega) \geq 0 \end{aligned}$$

$$\text{Thus, } (3 - 4\sin^2(k\omega) < 0 \Rightarrow sin^2(k\omega) \geq \frac{3}{4}$$

Both the condition contradicting with sin^2 values.Thus, this particular set of points can't be shattered with this ω.

It is not related to VC-dimension. Actually VC-dimension of sine function is infinite as per above question. There would be some ω which should shatter this points but for this ω this real number of point can't shatter.