Machine Learning from Data – IDC
HW6 – Theory

This assignment includes question related to learning theory.

1.

a. (10 pts) Let X be some infinite space of instances. Compute the VC-dimension of the following hypothesis space:

$$H = \{h: X \to \{-1, +1\}, |x: h(x) = -1| \leq 100\}$$

The hypothesis space contains hypotheses that can return -1, 100 times or less.

b. (10 pts) Give an example of an instance space X and a space H binary hypotheses on X, such that:

$$VC(H) = 2019$$

c. (20 pts) Consider the hypotheses space of all linear classifiers in the plain. That is, let $X = \mathbb{R}^2$ and then:

$$H = \left\{h: \exists\, w_1, w_2, b \in \mathbb{R}\ s.t\ h(x_1, x_2) = \begin{cases} +1 & w_1x_1 + w_2x_2 + b > 0 \\ -1 & w_1x_1 + w_2x_2 + b \leq 0 \end{cases}\right\}$$

Show that VC(H) = 3 by performing the following steps.

1) Find a set of size 3 that H shatters.
2) Show that no set of size 4, $A = (z_1, z_2, z_3, z_4), z_i \in \mathbb{R}$ can be shattered by H.

Guidance: First prove the following lemma:

Lemma 1: Suppose a linear classifier h obtains prediction $y \in \{-1, +1\}$ on a set of points $z, z' \in \mathbb{R}^2$ $(h(z) = h(z') = y)$. Then it also obtains the same prediction on any intermediate point. Namely,

$$\forall \alpha \in [0,1]\ \ h\big((1-\alpha)z + \alpha z'\big) = y$$

And use it in each of the following 3 possible cases:

a) The convex hull of A forms a line.
b) The convex hull of A forms a triangle.
c) The convex hull of A forms a quadrilateral.

d. (20 pts) Consider the hypotheses space of all linear classifiers in d dimensional Euclidean space. That is, let $X = \mathbb{R}^d$ and then:

$$H = \left\{h: \exists\, \overline{w} \in \mathbb{R}^d, b \in \mathbb{R}\ s.t\ h(\bar{x}) = \begin{cases} +1 & \overline{w}\bar{x} + b > 0 \\ -1 & \overline{w}\bar{x} + b \leq 0 \end{cases}\right\}$$

Show that VC(H) = d+1.

2. (20 pts) Let $X = \{0,1\}^n$ (all Boolean strings of length n). Let $C = H$ the set of all conjunctions of literals over X (e.g. $x_1 \wedge \neg x_4 \wedge x_n$ is in C and H). Define an algorithm L so that C is PAC-learnable by L using H. Prove all your steps.

3. (20 pts) Let $X = \mathbb{R}^2$. Let $C = H$ the set of all isosceles straight triangles with sides parallel to the axes and with their head vertex on the lower left (see picture). Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.