Theoretical Machine Learning - COS 511

Homework Assignment 5

Due Date: two weeks from announcement, in class

(1) Consulting other students from this course is allowed. In this case - clearly state whom you consulted with for each problem separately.

(2) Searching the internet or literature for solutions, other than the course lecture notes, is NOT allowed.

Ex. 1:
Prove that for $m \geq d$,

$$\sum_{i=0}^{d} \binom{m}{i} \leq \left(\frac{em}{d}\right)^d$$

Ex. 2:
In this exercise we consider the attribute set $X = \mathbb{R}^d$ for some $d \geq 1$, and the label set $Y = \{-1, 1\}$. In this classification problem, we consider the set of all hyperplanes as candidate hypotheses.

More accurately, we define a hyperplane $H \in \mathbb{R}^d$ as $H = \{x \mid a^T x = b\}$ for some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ (For example, in $\mathbb{R}^2$ a hyperplane is simply a line), and its corresponding hypothesis $h : \mathbb{R}^d \to \{-1, 1\}$ as a function $h(x) = \text{sign}(a^T x - b)$. We assume that $\text{sign}(0) = 1$. We denote by $\mathcal{H}$ the set of all hypotheses of this kind. Show that:

(1) For $d = 2$, the VC dimension of $\mathcal{H}$ is 3.

(2) For $d = 3$, the VC dimension of $\mathcal{H}$ is 4.

(3) For any value of $d$, it exists that $VC(\mathcal{H}) \geq d + 1$. I.e., show that there exists a set of $d + 1$ points that can be perfectly classified for any labelling.

(4) For any value of $d$, it exists that $VC(\mathcal{H}) \leq d + 1$. I.e., show that there does not exist a set of $d + 2$ points that can be perfectly classified for any labeling.
Ex. 3:
For \( m \geq d \), let \( \binom{m}{d} = \sum_{i=0}^{d} \binom{m}{i} \). Prove that

\[
\binom{m}{d} = \binom{m-1}{d} + \binom{m-1}{d-1}
\]

Ex. 4-7:
The following questions are taken from the book draft on online convex optimization (reading material number 1).

(1) problem 3 in chapter 6.
(2) problem 4 in chapter 6.
(3) problem 5 in chapter 6.
(4) problem 6 in chapter 6.