HW1 Intro to ML

---

1. Open ended task

2. (a) We are given the hypothesis $\mathcal{H} = \{h_i(x) = \text{sign}(x[i]) \mid i = 1, \ldots, d\}$ and $x \in \mathbb{R}^d$. To choose a hypothesis from $\mathcal{H}$, we can use the MAJORITY$_{\mathcal{H}}$, aka, HALVING learning rule. Under the realizability assumption, we see that

$$M = \log_2 |\mathcal{H}|$$

where in our case, $|\mathcal{H}| = d$ and therefore,

$$M = \log_2 d$$

(b) To demonstrate that the Nearest Neighbor (NN) learning rule can incur an exponential number of mistakes in d, consider a sequence of binary-coordinate points in $\{\pm 1\}^d$. This sequence is constructed such that for any given point, the nearest previously seen point has the opposite label, forcing a mistake by the NN rule.

Starting in $\mathbb{R}^3$, take $x_1 = (1, 1, 1)$ with label $y_1 = 1$. The next three points, $x_2 = (-1, 1, 1)$, $x_3 = (1, -1, 1)$, and $x_4 = (1, 1, -1)$, are the nearest neighbors to x_1 and are labeled $y_2 = y_3 = y_4 = -1$. Continuing this pattern, we label the next "closest set" of unlabelled points, which are closest to points labeled -1, with 1, and so on. Each point is labeled opposite to its nearest neighbor, resulting in a pattern that guarantees the NN rule makes a mistake at each step.

Generalizing to $\mathbb{R}^d$, we label the first point $x_1 = (1, \ldots, 1)$ as 1, then label the next d points (which differ with the initial point in exactly one coordinate) as -1, and alternate labels in subsequent layers of the d-cube. The NN rule will make a mistake for each of the $2^d - 1$ points following x_1.

Finally, to satisfy the realizability by H, we set the label to match the sign of the last coordinate. This results in a true hypothesis within H that correctly classifies all points, reducing the mistake count to $2^{d-1} - 1$, which is still exponential, $2^{\Omega(d)}$.