ELL880: Computational Learning Theory and the Mind

Minor Test I, Maximum marks: 20

Section 1.

1. Recall the two animal learning examples we discussed: the rats who learn to avoid poisnous baits, and the pigeons who develop the superstition that a certain activity leads to them being rewarded with food.

   (a) What did we infer from these examples about the kinds of principles a theory of (successful) learning should be aiming to formalise and enunciate? **[2]**

   (b) Recall that while the rats learn to associate the consumption of poisoned food with the nausea that follows it, they fail to learn associations between food and electric shock, or between sound and nausea. What does this observation specifically tell us about how we should seek to develop a theory of learning? **[1.5]**

2. Consider the papaya classification example we used in class. Suppose we have two real-valued features, colour (x_1) and softness (x_2), and a binary class label: tasty (1) / not tasty (0). We choose to apply the inductive bias of only considering the hypothesis class of *axis-aligned rectangles*. Formally, an axis-aligned rectangle is characterised by its x_1- and x_2-axis bounds, and given four real numbers a_1, a_2, b_1, b_2 such that $a_1 \leq a_2$ (x_1 bounds) and $b_1 \leq b_2$ (x_2 bounds), we can define the corresponding hypothesis as

$$h_{(a_1,a_2,b_1,b_2)}(x_1, x_2) = \begin{cases} 1 & \text{if } a_1 \leq x_1 \leq a_2 \text{ and } b_1 \leq x_2 \leq b_2; \\ 0 & \text{otherwise.} \end{cases}$$

   Our full hypothesis space then becomes:

$$\mathcal{H} = \{h_{(a_1,a_2,b_1,b_2)} : a_1 \leq a_2, \text{and } b_1 \leq b_2\}.$$

   (a) Suppose we define our learning algorithm (A) to be such that it returns the axis-aligned rectangle with the smallest area which encloses all the instances of tasty papayas in the training set S, sampled *i.i.d.* from the data-generating distribution $\mathcal{D}, f$ (assume realisability). Show that $A \in ERM_{\mathcal{H}}$. **[2.5]**

   (b) Now we want to show that A is a PAC learner for $\mathcal{H}$, under the realisability assumption. Towards this end, attempt the following steps.

      i. Under realisability, we must have $f \in \mathcal{H}$. So let $f = h_{(a_1^*,a_2^*,b_1^*,b_2^*)}$. Show that the rectangle corresponding to the hypothesis returned by A always lies inside of the rectangle corresponding to f. **[1]**

      ii. Let $a_1 \leq a_2^*$ be a number such that $h_1 = h_{(a_1,a_2^*,b_1^*,b_2^*)}$ corresponds to a rectangle whose probability mass under $\mathcal{D}$ is $\epsilon/4$, for some $\epsilon > 0$ (by choosing a_1 sufficiently close to a_2^*, we can make the rectangle sufficiently thin for this to happen). Similarly, let

      - $a_2 \geq a_1^*$ be a number such that $h_2 = h_{(a_1^*,a_2,b_1^*,b_2^*)}$ is a rectangle with probability mass $\epsilon/4$;
      - $b_1 \leq b_2^*$ be a number such that $h_3 = h_{(a_1^*,a_2^*,b_1,b_2^*)}$ is a rectangle with probability mass $\epsilon/4$; and
      - $b_2 \geq b_1^*$ be a number such that $h_4 = h_{(a_1^*,a_2^*,b_1^*,b_2)}$ is a rectangle with probability mass $\epsilon/4$.

      Show that if S happens to contain tasty papayas lying within *all* of the rectangles h_1, h_2, h_3, h_4, then the true loss of $h_S = A(S)$ is bounded by ϵ. **[5]**

      iii. So now we want the above condition (under which h_S is *approximately correct*) to hold with high ($1 - \delta$) probability, in order to establish PAC learnability. As a function of the number of training samples N, upper bound the probability that S does *not* contain a tasty papaya lying within the h_i rectangle, $\forall i \in \{1, 2, 3, 4\}$. **[3.5]**

iv. Obtain a union bound on the overall probability of the above condition which makes h_S approximately correct *not* holding. Note that this is the quantity we want to be bounded by δ, for some $\delta \in (0, 1)$. **[2.5]**

v. Use the above union bound to obtain the *minimal sample complexity*, as a function of ϵ and δ, which ensures that h_S is *probably approximately correct*. Hence conclude the proof of $\mathcal{H}$ being PAC learnable via A. **[2]**