CS60073: Advanced Machine Learning

Mid-semester Examination Spring 2018

Answer all THREE questions. Clearly write your steps of derivations. Answer all parts of a question together. Make suitable assumptions if necessary.

Time: 2 hrs. Total marks: 40.

---

1. Consider the hypothesis class of all decision lists $\mathcal{H}_{DL}$. The domain is the boolean hypercube $\{0,1\}^d$. The hypothesis class consists of all decision lists, which are a sequence of if-then-else rules of the form "*If* l_1 *then* b_1*, else if* l_2 *then* b_2*,..., else if* l_i *then* b_i*,..., else* b_k*.*", where l_i's are boolean literals (either x_i or $\overline{x}_i$ for some $i \in [d]$), and $b_i = \{0,1\}$.

(a) Design an algorithm for computing an ERM decision list from a training sample.

[5]

(b) Provide a bound on $|\mathcal{H}_{DL}|$, the hypothesis class of all decision lists. What upper bound on the sample complexity of PAC-learning decision lists does this imply?

[10]

2(a). Show that adding one function to a hypothesis class can increase VC-dimension (VCD) by at most one. That is, for a hypothesis class $\mathcal{H}$, and a function h over the same domain we have $VCD(\mathcal{H} \cup \{h\}) \leq VCD(\mathcal{H}) + 1$.

[10]

(b). Give an example of a class $\mathcal{H}$ over some domain X, and a function h, where the above bound is tight, i.e., $VCD(\mathcal{H} \cup \{h\}) = VCD(\mathcal{H}) + 1$.

[5]

3. Let us consider the boolean hypercube $\{0,1\}^n$. For a set $I \subseteq \{1,2,\ldots,n\}$, we define a parity function h_I as follows. On a binary vector $X = \{x_1, x_2, \ldots, x_n\} \in \{0,1\}^n$,

$$h_I(X) = \left(\sum_{i \in I} x_i \right) \mod 2$$

(That is, h_I computes parity of bits in I). What is the VC-dimension of the class of parity functions, $\mathcal{H}_{parity} = \{h_I : I \subseteq \{1,2,\ldots,n\}\}$?

[10]