CS671 - Machine Learning
Homework 1
Posted February 18, 2015
Due March 4, 2015

1. Let $\mathcal{X}$, the set of examples, be the set of natural numbers. The hypotheses space H consists of all intervals of the form $[a, b]$ with $a \leqslant b$. The concept that must be learned is an interval $[c, d]$, where all examples reside.

   (a) Let h_1, h_2 be two hypotheses. What does it mean in this context that h_1 is more specific than h_2?

   (b) Design an algorithm that learns the target concept.

2. Let $\mathcal{X}$ be a set of examples. Suppose that the hypotheses space consists of *all* functions $h : \mathcal{X} \longrightarrow \{\prime, \infty\}$. Prove that any unobserved example satisfies exactly half of hypotheses in the current version space, regardless of which training examples had been observed.

3. Consider a learning problem where each instance is described by a conjunction of n Boolean attributes $A_1, \ldots, A_n$. Here, a Boolean attribute is an attribute whose domain consists of two values, **t** and **f**. Thus, a typical instance would be

   $$(A_1 = \mathbf{t}) \wedge (A_2 = \mathbf{f}) \wedge \cdots \wedge (A_n = \mathbf{t}).$$

   Consider a hypothesis space H in which each hypothesis is a disjunction of constraints over these attributes. For example, a typical hypothesis would be

   $$(A_1 = \mathbf{t}) \vee (A_5 = \mathbf{f}) \vee (A_7 = \mathbf{t}).$$

   Design an algorithm that accepts a series of training examples and outputs a consistent hypothesis *if one exists*. Your algorithm should run in time that is polynomial in n (the number of attributes) and in m, the number of training examples.