Question 1.

Let h, h' be propositional conjunctions. Is $h' \models h$ equivalent to $h \subseteq h'$? Justify your answer.

Answer:

The two relations are not equivalent. For example $p \wedge \neg p \models q$ but $q \nsubseteq p \wedge \neg p$. The equivalence would hold if h' was not tautologically false.

Question 2.

Consider basis expansion introducing additional n attributes holding the inverted values contained in the original observation x. Such transformation turns general conjunctions (disjunctions) into *monotone* ones, thus into linearly separable ones, i.e., learnable by Winnow algorithm. How will the transformation change the algorithm's mistake bound?

Answer:

Winnow has the mistake bound

$$2 + 2k \log n.$$

It changes to

$$2 + 2k \lg 2n = 2 + 2k(1 + \lg n) = 2 + 2k \lg n + 2k$$

i.e., only by an additive constant $2k$.

Question 3.

Consider the halving algorithm with hypothesis class (initial hypothesis) $\mathcal{H}_1$ of all non-contradictory conjunctions on 3 propositional variables.

1. Determine $|\mathcal{H}_1|$.

   Answer:

   $3^3 = 27$ (Each of the 3 variables may be absent, positive, or negative in the conjunction.)

2. Give an upper bound on $|\mathcal{H}_2|$ given that first prediction was incorrect.

   Answer:

   The halving algorithm decides by majority vote so at least 14 hypotheses in $\mathcal{H}$ were inconsistent with the observation; those get deleted and at most 13 remain.

Question 4.

Consider halving algorithm with the initial version space $\mathcal{H}$ consisting

(a) of all conjunctions of exactly 3 different non-negative literals, i.e.,

$$\mathcal{H} = \{ p_i \wedge p_j \wedge p_k \mid 1 \leq i < j < k \leq n \}$$

(b) of all conjunctions that use some of the given variables (and the empty conjunction).

(c) of all n-CNFs.

1. For each scenario, determine if the learner learns $\mathcal{H}$ online (in the mistake-bound model) and justify your answer.

   Answer:

   The halving algorithm makes at most $\lg |\mathcal{H}|$ mistakes when learning a hypothesis from $\mathcal{H}$.

(a) $|\mathcal{H}| = \binom{n}{3} \leq n^3/3! = n^3/6$, so $\lg|\mathcal{H}| \leq 3\lg n \leq \text{poly}(n)$. Hence, the learner learns $\mathcal{H}$ online.

(b) $|\mathcal{H}| = 2^{2n}$, so $\lg|\mathcal{H}| = 2n$ is polynomial in n and the algorithm learns $\mathcal{H}$ online.

(c) $|\mathcal{H}| = 2^{\sum_{i=1}^{n}\binom{n}{i}2^i} = 2^{3^n-1}$, so $\lg|\mathcal{H}| = 3^n - 1$. Hence, the algorithm does not learn $\mathcal{H}$ online.

2. For each scenario where the learner successfully learns in the mistake bound model, decide if they learn efficiently as well. Assume that checking the consistency of a single hypothesis with an observation takes a unit of time.

Answer:

(a) $|\mathcal{H}| \leq \text{poly}(n)$, so yes.

(b) $|\mathcal{H}|$ is super-polynomial in n, hence no.

3. For the **first** case, assume the first example is $(0, 1, 1, 1, \ldots 1)$ and it is a negative instance. What will be the learner's prediction for the second example, which is $(0, 1, 0, 1, \ldots)$? Justify your answer.

Answer:

The prediction will be 0 (negative). On the first observation, all conjunctions not containing p_1 are deleted from the version space. Hence, all remaining conjunctions include p_1, which is assigned 0 by the second observation. All hypotheses in the version space are thus evaluated to $false$, hence the predicted label will 0.