15-251 : Great Theoretical Ideas In Computer Science

Fall 2013

Assignment 9: The Very Best Assignment

Due: Thursday, Nov. 14, 2013 11:59 PM

Name: ____________________

Andrew ID: ____________________

Question:	1	2	3	4	5	Total
Points:	25	25	25	25	10	110
Score:						

1. ## Pikachu

Hello there! Welcome to the world of CONCEPT CLASSES! My name is KRUSKAL! People call me the CONCEPT CLASS PROF! This world is inhabited by creatures known as CONCEPT CLASSES! For some people, CONCEPT CLASSES are pets. Others use them for fights. Myself...I study CONCEPT CLASSES as a profession. But there are a couple of things I haven't been able to figure out. Can you help me?

(5) (a) Show that a finite concept class C has VC dimension at most $\log_2 |C|$.

Solution:

(10) (b) Let C_1 and C_2 be two concept classes. Define the "intersection class"

$$C = \{c \mid \exists c_1 \in C_1, c_2 \in C_2 \text{ s.t. } \forall x \in X\ c(x) = + \iff c_1(x) = + \wedge c_2(x) = +\}.$$

That is, every concept $c \in C$ is the "intersection" of some concepts $c_1 \in C_1$ and $c_2 \in C_2$ such that for every point $x \in X$, c labels x as $+$ if and only if both c_1 and c_2 label x as $+$. Recall from lecture that for any set of examples S and any concept class C', $\pi_{C'}(S)$ is the number of ways of labeling examples in S using concepts from C'. Let $\pi_{C'}(m) = \max_{S,|S|=m} \pi_{C'}(S)$ denote the maximum number of labelings C' can achieve for any set of m training examples.

Show that $\pi_C(m) \leq \pi_{C_1}(m) \cdot \pi_{C_2}(m)$.

Solution:

(10) (c) In my youth, I was a serious Pokémon trainer. In particular, I had caught Pokémon of many types and labeled them by type (they were my training examples). I wanted to find a *consistent hypothesis* to reason about them. Back in those days, we didn't have new-fangled inventions like the Pokédex that would simply give me a consistent hypothesis. However, I had at my disposal a PAC learning algorithm itself. Recall the following from the lecture.

Finding a consistent hypothesis: Given a set S of labelled points, find a hypothesis that is consistent with S, that is, it labels every point in S correctly.

A PAC learing algorithm: Given $\epsilon > 0$, $\delta > 0$, and $m_0(\epsilon, \delta)$ points from some distribution D, it outputs a hypothesis such that the error of the hypothesis w.r.t. D is at most ϵ with probability at least $1 - \delta$.

In lecture, we saw how to use an algorithm for finding a consistent hypothesis to construct a PAC learning algorithm: Just find a hypothesis consistent with $m_0(\epsilon, \delta)$ points sampled from the input distribution. However, returning a consistent hypothesis is just one way of PAC learning; not all PAC learning algorithms output a hypothesis consistent with given examples. Your task is to achieve the direction opposite to what was done in the lecture.

Given a set S of labelled points, $\delta > 0$, and an arbitrary PAC learning algorithm A that may not always output a consistent hypothesis, show how you can use A to find a hypothesis that is consistent with S with probability at least $1 - \delta$.

[Hint: The PAC learning algorithm would require samples from some distribution D', an accuracy parameter ϵ', and a confidence parameter δ'. Given S, choose D', ϵ', and δ' appropriately so that the PAC learning algorithm returns a hypothesis that is consistent with S with probability at least $1 - \delta$.]

Solution:

2. Dugtrio

(10) (a) Later on in life, I became interested in triangles of types that form a 3-cycle of super-effectiveness, such as Grass, Fire, Water or Dark, Fighting, Psychic. To help me become better at studying type triangles, I decided to study actual triangles on the real plane.

In particular, I have some points on the real plane. I use concepts from a concept class TRIANGLE to label them. Each concept in TRIANGLE, as you may have guessed, is a triangle that labels all points inside it as $+$ and all points outside it as $-$. Prove that the VC-dimension of TRIANGLE is at least 7 by giving (and proving) an example of 7 points that can be shattered. [Hint: Try to argue why the vertices of a regular heptagon can be shattered.]

Solution:

(15) (b) Prove that the VC-dimension of TRIANGLE is less than 8 (and thus equals 7) by showing that no set of 8 points can be shattered.
[Hint: Consider two cases - when one of the vertex is internal (inside the convex hull of the others), and when none of the vertices is internal.]

Solution:

3. Scyther

After getting bored with 3-cycles, I decided to look for something bigger. In particular, I was looking for a chain containing all of the Pokémon types such that each type in the chain was super effective against the next type in the chain, but the last type in the chain was also super effective against the first type. I eventually realized that looking for a Hamiltonian cycle in the graph of the types was exactly what I wanted, but unfortunately this was still difficult. One day, my grandson Gary came to me excitedly claiming that he had found such a Hamiltonian cycle, but I was skeptical. So we decided that he would prove that a Hamiltonian cycle in G exists using the following interactive proof system. Gary is the prover, Kruskal is the verifier.

1. At the beginning of each round, Gary creates H, an isomorphic graph to G.
2. Gary commits to H using a cryptographic commitment scheme (so he can't change H later but Kruskal has no information about H).
3. Kruskal then randomly chooses one of two questions to ask Gary. He can either ask him to show the isomorphism between H and G, or he can ask him to show a Hamiltonian cycle in H.
4. If Gary is asked to show that the two graphs are isomorphic, he first uncovers all of H and then provides the vertex translations that map G to H. Kruskal can verify that they are indeed isomorphic.
5. If Gary is asked to prove that he knows a Hamiltonian cycle in H, he translates his Hamiltonian cycle in G onto H using the isomorphism between G and H known to him, and only uncovers the edges on the Hamiltonian cycle in H. Kruskal can verify that this is indeed a Hamiltonian cycle.

They repeat this some number of times. FYI, this is actually a zero-knowledge proof. Try to reason about it for fun, although you are not required to prove that :-)

(25) (a) Show that this protocol is sound and complete.

Solution:

4. Sudokuwoodo

Growing tired of Pokémon in my old age, I've found a new love: Sudoku! Of course, wanting to impress me, Gary comes up to me every day bragging about the complicated Sudoku boards that he's solved. I'm as skeptical of his Sudoku-solving abilities as I am of his cycle-finding abilities, so I want to test his solution, but I don't want to have the puzzle spoiled before I can solve it myself. So I decided to test his solution using the following interactive proof system. Again, Gary is the prover, and Kruskal is the verifier.

1. First, Gary chooses a random permutation of the numbers 1-9 and permutes his solution using this.
2. Next, he commits all the squares in this solution, so Gary can no longer edit them but Kruskal can't see them.
3. Kruskal can then choose to see one of 28 choices.
   - Any of the 9 rows of the board
   - Any of the 9 columns of the board
   - Any of the 9 subgrids of the board
   - Just the squares of the original puzzle with Gary's permutation applied
4. Gary then reveals the appropriate squares to Kruskal. If it's a row, column, or subgrid, Kruskal accepts iff each number from 1-9 appears exactly once. If it's the original puzzle, Kruskal accepts iff Gary presents a valid permutation of the original puzzle.

As usual, they repeat this protocol some number of times. Once again, try to reason about zero-knowledge of this protocol for fun, but you're not required to show that.

(25) (a) Prove that this protocol is sound and complete.

Solution:

5. Bonus: Voltorb

(10) (a) At another time, I decided to study the mechanics of the Pokéball. It turns out that the VC-dimension of the Pokéball is essential to its capturing ability, so I wanted to discover that.

In particular, we have points in $\mathbb{R}^3$. We use a concept class SPHERES, such that each concept in this class, as the unimaginative name suggests, is a sphere that labels all points on or inside the sphere as $+$ and all points outside the sphere as $-$. Show that the VC-dimension of SPHERES is 4.
[Hint: For an upper bound, show that if SPHERES can shatter a training set, then so can PLANES. Each concept in PLANES is a plane in $\mathbb{R}^3$ that labels all points on the plane or on a fixed side of the plane as $+$, and all points on the other side of the plane as $-$. You are given that the VC-dimension of PLANES is 4.]

Solution: