CS 5350/6350, DS 4350: Machine Learning, Fall 2024

Sample Midterm Questions

This document contains a set of questions to give a flavor of the midterm exam. (The actual midterm will not be as long as this.) Feel free to discuss these questions with the instructor, the TAs and other students.

1. How would you train a decision tree using the ID3 algorithm if some attributes are missing?
2. Step through the process of constructing a decision tree using the ID3 algorithm for a small dataset like the Tennis data in the lecture.
3. Show that Dataset 1 in table 1 is linearly separable by providing a linear threshold unit that correctly classifies the examples.
4. How would you avoid overfitting when you use the decision tree algorithm? Why might shorter decision trees be more robust to noise in the training data?
5. Consider Dataset 2 in table 2 and answer the following questions
   (a) Which of the three features x_1, x_2 or x_3 has the highest information gain?
   (b) Construct a decision tree of depth one (i.e. that uses just one feature) using the feature with the highest information gain. Justify your choice for the labels on the leaves.
   (c) What is the training error of the tree you constructed for the previous question?
6. For each function below, state whether it can be written as a linear threshold unit in terms of the variables specified. If it can be written as one, write the linear threshold unit that is equivalent to the function. If not, suggest a transformation of the underlying space so that the function is linear in the new space.
   (a) $\neg x_1$
   (b) $x_1 \vee \neg x_2$
   (c) $(x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_3)$

x_1	x_2	x_3	y
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0

Table 1: Dataset 1

x_1	x_2	x_3	y
-1	-1	-1	1
-1	-1	1	1
-1	1	-1	-1
-1	1	1	1
1	-1	-1	-1
1	-1	1	-1

Table 2: Dataset 2

7. Show that the Halving algorithm for a finite concept space C will not make more than $\log |C|$ mistakes. Apply this to get a limit on the number of mistakes the algorithm will make for the class of k-conjunctions of n Boolean variables.

8. State with an explanation whether the following are true or false.

   (a) The mistake bound model assumes that training and test examples are drawn from the same fixed, but unknown distribution.

   (b) The Perceptron mistake bound theorem guarantees that the algorithm will find a linear separator for *any* dataset.

   (c) A learning algorithm that makes a finite number of mistakes on any dataset is called a mistake bound algorithm.

9. Prove the Perceptron mistake bound.

10. Using Dataset 2 in table 2, step through the Perceptron algorithm, starting with all weights and the bias term being zero.

11. Prove a mistake bound for the margin Perceptron. More formally, the margin Perceptron updates its weights for an example $\mathbf{x}_i$ with label y_i if $y_i \mathbf{w}_t^T \mathbf{x}_i \leq \mu$. Here, μ is a fixed parameter.

    As with the standard Perceptron, suppose all examples are contained in a ball of radius R and let a unit vector $\mathbf{u}$ perfectly classify the data with margin γ.

    (For such a proof, you will have to follow the template of the Perceptron mistake bound proof: First, prove that $\mathbf{u}^T \mathbf{w}_t$ keeps increasing with each update Then, prove that $\|\mathbf{w}_t\|$ is bounded above. Finally, combine these two bounds in exactly the same fashion as in the Perceptron case to get an inequality involving the number of updates.)

12. How many mistakes will the Perceptron algorithm make for disjunctions with n attributes? To answer this, you will first have to identify what R and γ are for this concept class. To get started with γ, see what happens when $n = 2$.

13. Suppose our learning problem has n binary features. What is the size of the hypothesis space consisting of all decision trees over this space?

14. You wish to learn a hidden concept f using m training examples that are drawn from a distribution D. If the training set is called S and the hypothesis that your learning generates is h, write expressions for the empirical and true errors.