CS 6375
Fall 2018
Midterm
10/10/2018

Name (Print): ______________________________

This exam contains 10 pages (including this cover page) and 5 problems. Check to see if any pages are missing. Enter all requested information on the top of this page, and put your initials on the top of every page, in case the pages become separated.

You may **NOT** use books, notes, or any electronic devices on this exam. Examinees found to be using any materials other than a pen or pencil will receive a zero on the exam and face possible disciplinary action.

The following rules apply:

- **Organize your work**, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
- **To ensure maximum credit** on short answer / algorithmic questions, be sure to **EXPLAIN** your solution.
- **Problems/subproblems** are not necessarily ordered by difficulty. Be sure to read each of the questions carefully.
- **Do not** write in the table to the right.

Problem	Points	Score
1	24	
2	25	
3	15	
4	31	
5	5	
Total:	100	

1. (24 points) **True or False and Explain:** For each of the following statements indicate whether or not they are true or false and explain your reasoning. Simply writing true or false without correct reasoning will receive no credit.

   (a) The union bound for $p(A \cup B \cup C)$ over events A, B, C is exact when the events are independent.

   (b) The VC dimension of a finite hypothesis space, H, is at least $\log |H|$.

   (c) Consider two lines in $\mathbb{R}^2$ determined by weights $w_1, w_2 \in \mathbb{R}$ and biases $b_1, b_2 \in \mathbb{R}$ respectively. Given training data for a linear regression problem, if $f(w, b)$ computes the squared loss of $w^T x + b$ over the training data set for each w and b, then $f(\frac{w_1+w_2}{2}, \frac{b_1+b_2}{2}) \leq \max\{f(w_1, b_1), f(w_2, b_2)\}$.

(True or False continued)

(d) Consider forming the dual optimization problem (using the method of Lagrange multipliers) of a constrained optimization problem. If you form the dual optimization problem (again using the method of Lagrange multipliers) of the dual optimization problem, you recover the primal problem, i.e., is the dual of the dual equal to the primal?

(e) Suppose two hypotheses have the same accuracy on a given training data set. The hypothesis with higher accuracy on a held-out validation set will also have higher accuracy on a held-out test set.

(f) Bagging is likely to be more useful than boosting if the hypothesis space is very expressive.

2. **Fixed Squares:** Consider a binary classification problem with data points in $\mathbb{R}^2$. Fix k distinct centers, $c_1, \ldots, c_k \in \mathbb{R}^2$. Let H be the hypothesis space whose elements consist of k-tuples of axis-aligned squares $(S_1, \ldots, S_k)$ such that S_a is a square centered at the point c_a for each $a \in \{1, \ldots, k\}$. Note that the centers have been fixed in advance and only the sizes of the individual squares are permitted to vary. A point is classified as a plus by the k-tuple $(S_1, \ldots, S_k)$ if it is contained in at least one square and is classified as a minus otherwise.

   (a) (15 points) What is the VC dimension of this hypothesis space on $\mathbb{R}^2$ with $k = 3$? Prove it.

(Fixed Squares continued)

(b) (5 points) How would you define the margin of a classifier in this space?

(c) (5 points) Does the VC dimension increase, decrease, or remain the same if the squares are replaced with axis-aligned rectangles, which are also centered at the respective points?

3. **Almost-linear regression:** Consider regression task that fits a piecewise linear function of the form

$$f(x) = \begin{cases} a_1 x + b_1 & \text{if } x < 0 \\ a_2 x + b_2 & \text{if } x \geq 0 \end{cases},$$

where $a_1, a_2, b_1, b_2 \in \mathbb{R}$.

(a) (10 points) Given data points $(x^{(1)}, y^{(1)}), \ldots, (x^{(M)}, y^{(M)})$, formulate the regression problem to predict the y's as a loss minimization problem and explain how to apply gradient descent to find a global optimum.

(b) (5 points) Suppose that you wanted f to be continuous. Formulate the regression problem under this new constraint as a convex optimization problem.

4. **Relative SVMs:** Consider a binary classification problem for vectors in $\mathbb{R}^n$ using linear separators. For this problem, consider linearly separable, labeled training data of the form $(x^{(1)}, y^{(1)}), \ldots, x^{(M)}, y^{(M)})$, where $x^{(m)} \in \mathbb{R}^n$ and $y^{(m)} \in \{+1, -1\}$.

   (a) (5 points) Given a max-margin linear separator of the form $w^T x + b$ for the above data points, what is the distance of the point $x^{(m)}$ to the linear separator as a function of w and b?

   (b) (5 points) Express the constraint that "$x^{(m)}$ cannot be farther away from the linear separator than R times the size of the margin" as a pair of linear inequalities, for some constant $R > 0$.

(c) (15 points) Add the constraints from part (b) for each $x^{(m)}$ to the standard SVM objective without slack. Construct a dual of this optimization problem, treating R as a hyperparameter, using the method of Lagrange multipliers.

(d) (3 points) Can the kernel trick still be used in this dual formulation?

(e) (3 points) Explain the effect that the choice of R has on the solution to the modified SVM problem. How should you pick it in practice?

5. (5 points) **Short Answer:** Suppose that you are a professor teaching a graduate machine learning course. Your goal is to assign each student in the course one of five possible labels $\{A, B, C, D, F\}$ commensurate with their knowledge of machine learning as measured by their scores on individual homeworks and exams. Using your knowledge of decision trees, explain why you should give challenging exams rather than simple exams.