Name and CU email address: ______________________________

Exam 1
Fall 2019

CSCI 4622: Machine Learning
Instructor: C. Monteleoni

This exam has 6 questions, for a total of 30 points.

Definitions:

- Euclidean norm (i.e., L2 norm) of a vector $x \in \mathbb{R}^d$. Note: when there is no subscript to the norm, we will assume it is L2.

$$\|x\| = \|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$$

For grading. Please do not write here:

Question	Points	Bonus Points	Score
1	2	0	
2	1	0	
3	3	0	
4	12	0	
5	7	0	
6	5	0	
Total:	30	0	

1. (2 points) Total for Question 1: 2

Consider the following confusion matrix obtained by using a classifier on a test data set, to classify whether an image does or does not contain a cat. Which **ONE** of the following statements is **FALSE**.

	cat	no cat
cat	13	7
no cat	5	15

A. 13 images containing cats were accurately classified.

B. 5 images without cats were incorrectly classified.

C. This test data set contains 20 images.

D. The accuracy of the classifier on this test data set is 0.70.

1. **C**

Solution: The number of examples in the data set is the sum of all values in the confusion matrix. In this case there were 40 examples in the data set.

2. (1 point) Total for Question 2: 1

Which **ONE** of the following is **NOT** an ensemble method:

A. Bagging

B. Decision Tree

C. Random Forest

D. Boosting

2. **B**

3. Total for Question 3: 3

Answer the questions below, for the case of binary classification.

(a) (1 point) How many possible labelings are there of n points?

(a) 2^n

(b) (1 point) Consider a hypothesis class H. Denote as $D < \infty$, the VC-dimension of H. In big-O notation, state the upper bound provided by Sauer's Lemma, on the number of possible labelings of n points by classifiers in H.

(b) $O(n^D)$

(c) (1 point) Under what condition below is the bound in Part (b) asymptotically tighter (smaller) than your answer to Part (a)?

A. When $n > D$

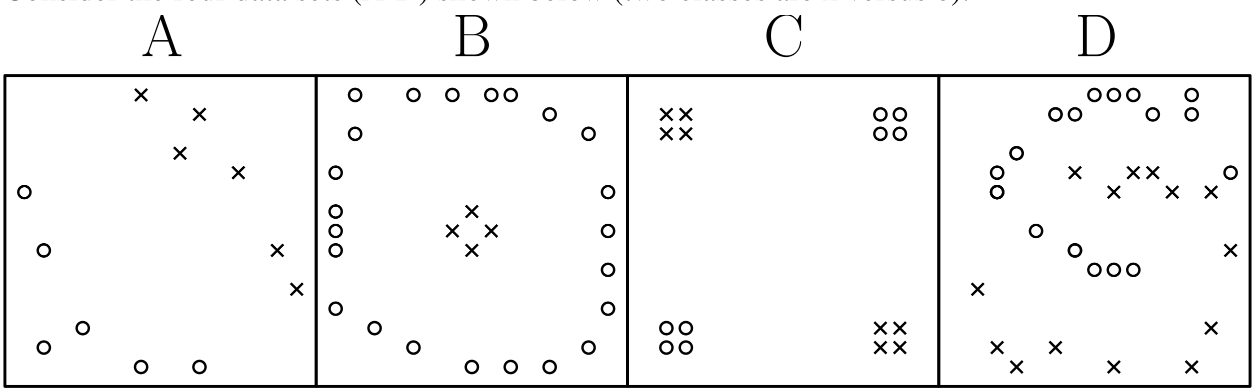
B. When $D > n$

C. Never

(c) **A**

4. Total for Question 4: 12

Consider the four data sets (A-D) shown below (two classes are x versus o):



For each of the following algorithms, on which of the datasets above would this algorithm get zero training error? Your answer will be the letter(s) above the data set(s) for which this holds, or "none."

(a) (4 points) Perceptron, when run for as many iterations as possible.

(b) (4 points) Depth two decision tree (i.e., two questions along any branch) that can ask threshold questions about a single feature (i.e., is $x_1 \geq 0.5$ and similar).

(c) (4 points) One-Nearest Neighbor classifier (k-NN where $k = 1$) with leave-one-out cross-validation evaluation of training error.

Solution:

(a) A

(b) A, C

(c) A, B, C

5. Total for Question 5: 7

For each of the following statements, check the box indicating whether the statement is True or False.

(a) (1 point) It is always good to drive training error as low as possible.
☐ True $\surd$ **False**

(b) (1 point) On a dataset with only binary features, the deepest a decision tree could possibly grow is D (where D is the number of features).
$\surd$ **True** ☐ False

(c) (1 point) The more complex a classifier f is, the more likely it is to generalize well to data that was not in the training set.
☐ True $\surd$ **False**

(d) (1 point) The goal of classification is to learn a function $f : X \to Y$ where X is some input space and $Y = \mathbb{R}$.
☐ True $\surd$ **False**

(e) (1 point) The VC-dimension of k-nearest-neighbor classifiers is not a function of k.
$\surd$ **True** ☐ False

(f) (1 point) The VC-dimension of linear classifiers in $\mathbb{R}^d$ is not a function of d.
☐ True $\surd$ **False**

(g) (1 point) In most cases, when a model is overfit, its training error is greater than its true test error.
☐ True $\surd$ **False**

6. Total for Question 6: 5

The following questions refer to the general template for a Boosting algorithm.

(a) (1 point) During an iteration of Boosting, once a classifier h_t has been learned, the weights on the training data are updated to _________ the weights on points which h_t classifies incorrectly.

Check the box which correctly fills in the blank:

$\surd$ *increase*

☐ *decrease*

(b) (4 points) The sequence of classifiers, h_t, learned in boosting are *weak classifiers*. With respect to the current weighting of the training data, check the boxes of **all** allowable error rates for a weak classifier.

$\surd$ **15%**

$\surd$ **49%**

☐ 51%

☐ 85%