Sample Midterm

Due: N/A

There are six problems, each worth 20% of the exam. The problem with the lowest score will be treated as an extra credit problem, worth 1/4 of a regular problem.

You may use a calculator, but no notes or textbook. Accessing a phone or any other device besides a standalone calculator will be considered an act of academic dishonesty.

Do not turn this page until we give the signal.

Banner ID: __________________________
Problem 1 (Representations)
Consider the following training set with two classes in $\mathbb{R}^2$. Draw a decision tree that perfectly fits this training data. Make sure to completely specify all branches, internal nodes, and leaves.
Problem 2 (Loss Functions)

Consider the 1-dimension hinge loss for a single training example \( x = 1 \) and \( y = 1 \) and a homogeneous halfspace with one weight \( w \in \mathbb{R} \):

\[
\ell_{\text{hinge}}(w) = \begin{cases} 
0 & \text{if } w \geq 1 \\
1 - w & \text{if } w < 1 
\end{cases} = \max\{0, 1 - w\}
\]

\( \ell_{\text{hinge}}(w) \) is a convex function that upper bounds another loss function called the ramp loss:

\[
\ell_{\text{ramp}}(w) = \begin{cases} 
0 & \text{if } w \geq 1 \\
1 - w & \text{if } 0 < w < 1 \\
1 & \text{if } w \leq 0 
\end{cases}
\]

a. Prove that \( \ell_{\text{ramp}}(w) \) is a non-convex function.

b. When learning a halfspace over many training examples, what could be an advantage of using the hinge loss over the ramp loss, and vice versa, what could be an advantage of using the ramp loss over the hinge loss?
Problem 3 (Optimizers)

Consider the hypothesis class of two-dimensional thresholds, $H = \{h_{a,b} : a \in \mathbb{R} \text{ and } b \in \mathbb{R}\}$ where:

$$h_{a,b}(x) = \begin{cases} 
1 & \text{if } x_1 \leq a \text{ and } x_2 \leq b \\
-1 & \text{otherwise}
\end{cases}$$

and $\mathcal{X} = \mathbb{R}^2$ and $\mathcal{Y} = \{-1, 1\}$.

Describe an algorithm for computing the ERM for this class in the realizable case. (You can assume 0-1 loss, although the solution will be the same for any reasonable loss function.) State the computational complexity of the algorithm in the context of a training data set of size $m$. 
Problem 4 (Empirical and Expected Risk)

For this problem, we are looking for responses that both indicate your assessment as to a possible accuracy change and your understanding of the algorithm that led to this assessment. Answers should be two or three sentences long and focus on the relevant and important issue.

a. We have trained a logistic regression model (binary vector input, binary label, no regularization) on a data set. Then, we create a new data set that is identical to the original but it includes a new feature that is set uniformly at random, with no strong correlation to any of the other features or the label, and run the same learning algorithm again. What would you expect to happen to the training and testing losses of the new learned model?

b. We have trained a logistic regression model (binary vector input, binary label, no regularization) on a data set. Then, we create a new data set that is identical to the original but includes a new attribute that is the Boolean negation of the label and run the same learning algorithm again. What would you expect to happen to the training and testing losses of the new learned model?
Problem 5 (Model Selection)
Consider the following (partially labeled) model selection curve for boosted halfspace classifiers learned with AdaBoost:

![Model Selection Curve](image)

a. Describe the likely interpretation of the following parts of the above figure, based on the bias-complexity tradeoff. Include a specific statement of what that part of the figure represents, and provide a brief explanation justifying your interpretation.

The horizontal axis (with values 1 through 7):

Curve A (solid line):

Curve B (dotted line):

b. If you were using the above model selection curve to choose a specific value on the horizontal axis to use for the corresponding task, which would you choose? Why?
Problem 6 (VC dimension)

Consider (again, see problem 3) the hypothesis class of two-dimensional thresholds, \( H = \{ h_{a,b} : a \in \mathbb{R} \text{ and } b \in \mathbb{R} \} \) where:

\[
h_{a,b}(x) = \begin{cases} 
1 & \text{if } x_1 \leq a \text{ and } x_2 \leq b \\
-1 & \text{otherwise}
\end{cases}
\]

and \( \mathcal{X} = \mathbb{R}^2 \) and \( \mathcal{Y} = \{-1, 1\} \).

What is the VC dimension of this hypothesis class? Provide a complete proof.