622 Exam 2

This is a preview of the published version of the quiz

Started: Jan 20 at 5:35pm

Quiz Instructions

This is the second midterm. It covers material post midterm 1. You have 2 hours from the time that you start the exam to submit. You must show your work on this exam. You are welcome to use a calculator, but you cannot use solvers (think eigenvalue solvers).

Question 1 — 5 pts

How can an SVM be used to solve the XOR problem? By "solve" we mean find a boundary that perfectly classifies the data.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 2

5 pts

True/False: PCA can be applied before all of the other algorithms we have learned in class. Explain.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 3

5 pts

True/False: A soft-margin SVM will always give the same results assuming the training data is the same. Briefly explain.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 4

5 pts

If I am going to apply PCA to the following data, what size with my covariance matrix be? (s_i stands for sample i)

$s_1 = (0, 1, 2, 1)$

$s_2 = (1, 3, 0, 2)$

$s_3 = (3, 5, 1, 1)$

$s_4 = (2, 2, 1, 4)$

$s_5 = (10, 1, 3, 5)$

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 5

5 pts

In a setting where there is no human oversight, is it more important to have high precision or high recall? Explain.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 6

5 pts

Assume you have an image of size $N \times N$. You are training a CNN with many layers. You only apply one filter at each layer and that filter always has a size of $K \times K$. How many layers would it take before the resulting feature map would be too small to apply a $K \times K$ filter? Assume there is no padding and you have a stride of 1. Give an expression for the number of layers using these variables (call this L).

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 7

5 pts

I have a large dataset. Should I separate the data into Train, Val & Test? Or should I use 5-fold Cross Validation? Explain why you chose one over the other. Would your answer change if I had a small dataset? Why or why not?

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 8

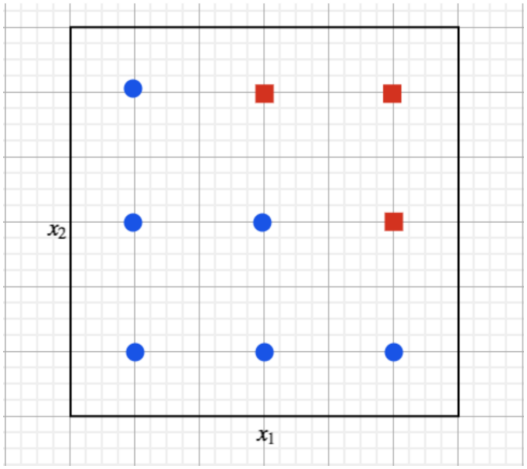
10 pts

We want to perform Adaboost to train a strong classifier using many weak classifiers. In this case, our weak classifiers are going to be axis-parallel lines (single lines that are either vertical or horizontal) that minimize the weighted training error. Everything on one side of the line will be classified as positive and on the other side will be negative. There are three parts to this question, so make sure to answer all three.

a) Using the dataset below, draw the decision boundary learned by the first classifier in the Adaboost algorithm.

b) Circle the point(s) with the highest weights at the beginning of the second iteration of Adaboost.

c) Draw the decision boundary learned by the the second classifier in the Adaboost Algorithm.



Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 9

15 pts

Given the following kernel, find the function $\phi(x)$ such that $K(x, z) = \phi(x) \cdot \phi(z)$. Each sample ($x$ and z) is one dimensional (they have one feature).

$$K(x, z) = \left(1 + xz + x^2 z^2\right)^2$$

Edit View Insert Format Tools Table

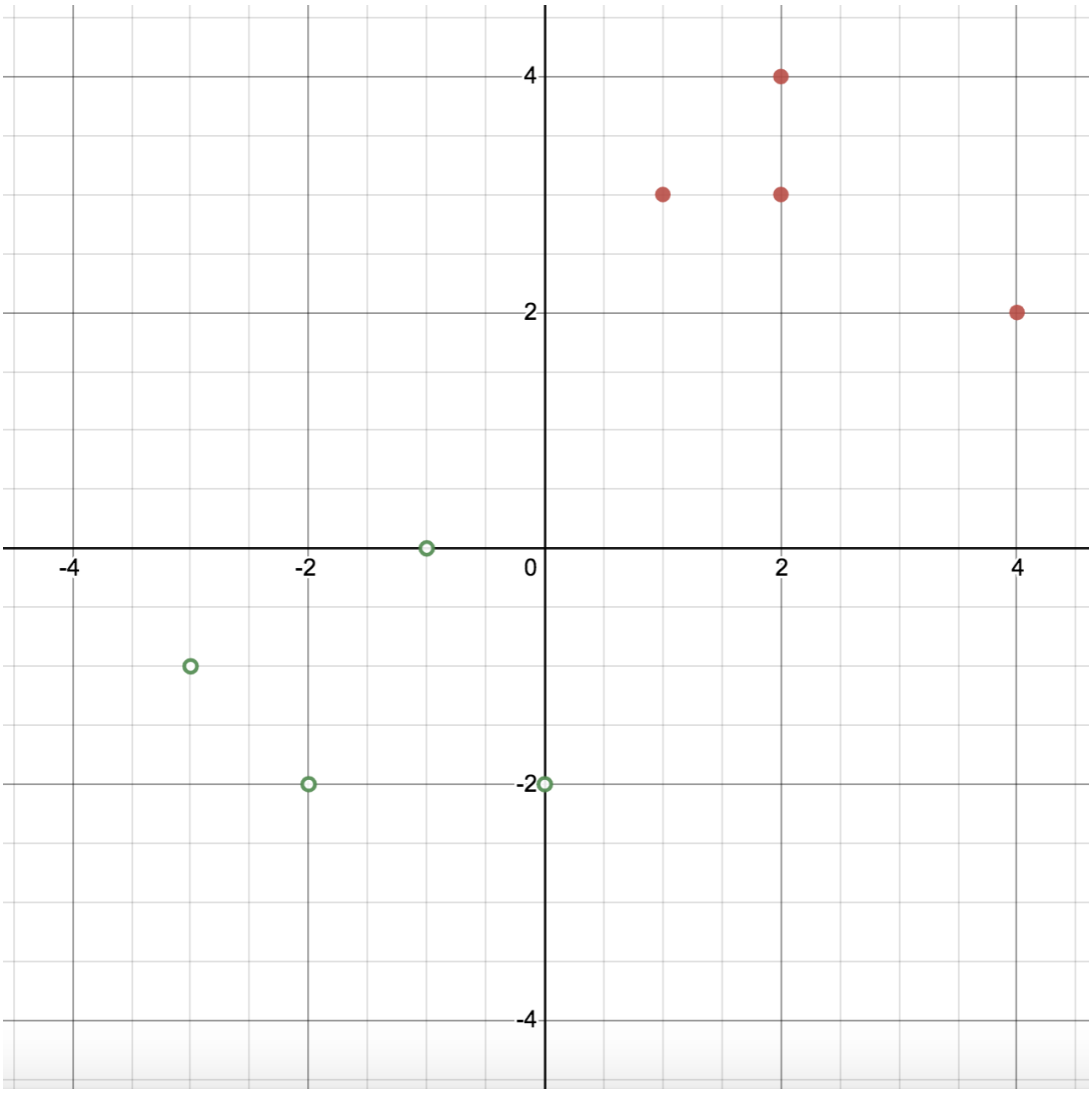
12pt Paragraph

p 0 words

Question 10

20 pts

Given the following data, what w and b would a hard-margin SVM return? Red filled circles are the negative class. Green empty circles are the positive class. Assume all points lie on integer coordinates.



Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

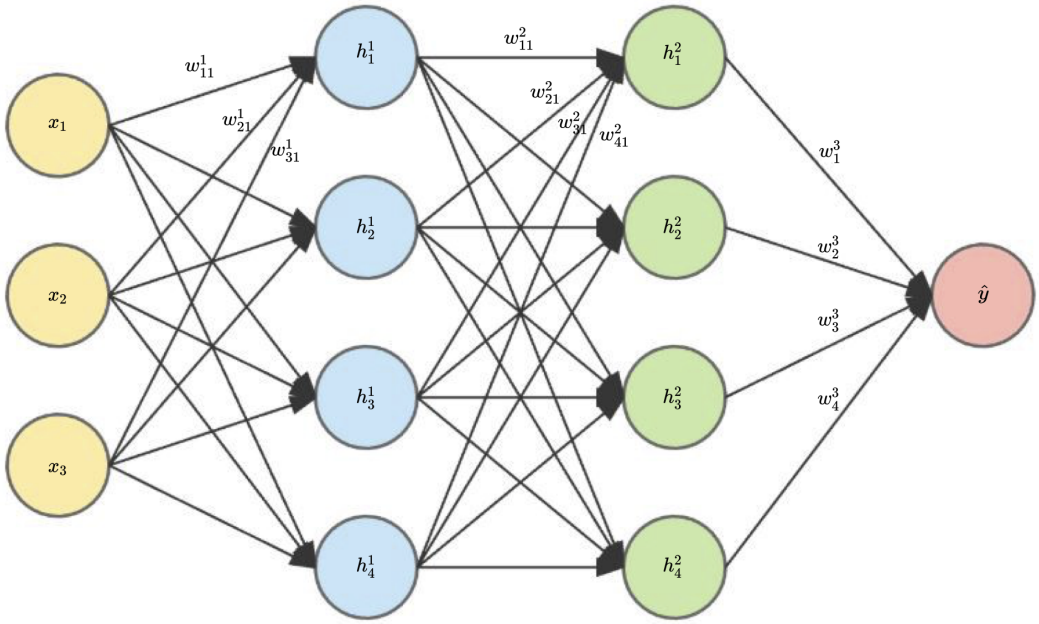
Question 11

20 pts

Given the following neural network, give the expression for $\frac{\partial L}{\partial w_{11}^1}$. Assume $L = (y - \hat{y})^2$, $h_i^1 = e^{w_{i1}^1 \cdot x}$ (that is, the first hidden layer applies e to its input), and there is no nonlinearity at the second hidden layer nor at the output.

Remember the notation for w_{ij}^k means the k^{th} layer of weights going from the i^{th} node to the j^{th} node.

Write out the expression for $\frac{\partial L}{\partial w_{11}^1}$ using partial derivatives (e.g. $\frac{\partial L}{\partial w_{11}^1} = \frac{\partial L}{\partial something} \frac{\partial something}{\partial somethingelse} \cdots$) and then identify each partial derivative separately (e.g. $\frac{\partial L}{\partial something} = blahblahblah$, $\frac{\partial something}{\partial somethingelse} = yadayada$). This is the only way you will receive partial credit!!!



Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Not saved Submit Quiz