

# CS 412 Final

May 6th

Name: \_\_\_\_\_

UIC NetID: \_\_\_\_\_

UIN: \_\_\_\_\_

**Do not start the exam until told to do so**

There is extra space on the back of this page if you need more room. If you put an answer on these extra sheets, indicate clearly on the question page where your solution can be found. This exam will be 120 minutes in length. Please budget your time well so that you can show everything that you know.

Question	Max Points	Earned Points
1	30	
2	40	
3	20	
4	20	
5	20	
6	20	
<b>Total</b>	<b>150</b>	

- These questions may not necessarily be in order of difficulty, skip around if you get stuck
- Intermediate steps may not always be necessary, but they can earn you partial credit if your solution is incorrect
- You will also receive an appendix page with all of the formulae you will need for the exam
- If you need any clarification, please ask
- Good luck!

This page intentionally left blank

# 1 Multiple Choice

Answer the following multiple choice/fill in the blank questions. For all eight multiple choice, circle the best answer. For fill in the blank, you may write your answer below the question.

1) With decision trees, post-pruning is a method which does what?

- (a) Reduces underfitting
- (b) Reduces overfitting
- (c) Reduces computation time
- (d) Increases accuracy

2) For what reason is it important to only run on the test once?

- (a) Reducing runtime
- (b) Variance will increase
- (c) The test error will not reflect the actual error
- (d) The test set is usually too small

3) Which of the following is **not** an ensemble method?

- (a) Random Forest
- (b) AdaBoost
- (c) kNN (where  $k \geq 2$ )
- (d) Stacking

4) Which of the following are used to draw an ROC curve?

- (a) Precision and Recall
- (b) Sensitivity and Specificity
- (c) Bias and Variance
- (d) None of the above

- 5) What was the major problem that makes deep learning necessary?
- (a) Backpropagation fails with too many layers
  - (b) There are too few features
  - (c) Computation times were too long
  - (d) Variance was too high
- 6) What is the cause of most ethical problems in ML? (As per lecture)
- (a) Negligence
  - (b) Malice
  - (c) Underfitting
  - (d) All of the above
- 7) A polynomial degree 4 kernel on two features has how many terms? Include 1.
- (a) 5
  - (b) 8
  - (c) 10
  - (d) None of the above
- 8) Which of the following cannot create linear decision boundaries?
- (a) Logistic Regression
  - (b) SVM
  - (c) Neural Networks
  - (d) Naive Bayes
- 9) Which of the following is an appropriate description for cost sensitive learning ?
- (a) Some classes are very rare in the data set
  - (b) When there is a different model for different regions in the feature space
  - (c) There are too many features to learn efficiently
  - (d) The misclassification error as asymmetric

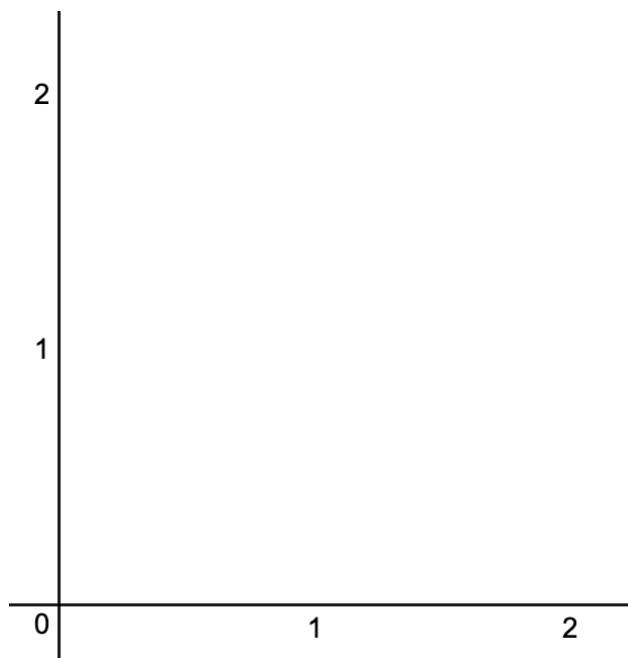
- 10) \_\_\_\_\_ is a form of classical statistical inference while \_\_\_\_\_ is a form of Bayesian statistical inference.
- 11) \_\_\_\_\_ is the error penalty parameter in SVM
- 12) A reinforcement learning approach is called \_\_\_\_\_ if it does not try to understand how actions lead to rewards or changes to the environment
- 13) A reinforcement learning method which does not choose the moves is called \_\_\_\_\_
- 14) The Gauss-Markov theorem states that the linear regression solution with minimal error also has minimal \_\_\_\_\_ .

## 2 Short Answer

a) What is the difference between L1 and L2 regularization?

b) What is unsupervised learning?

- c) Draw a dataset of Xs and Os which cannot be decided by NN with two internal nodes in one layer.



- d) What does it mean for a model to be unstable? Give an example of an unstable model

e) Give an example of an ethical problem in ML that violates the idea of consent

f) What is deep learning and why is it difficult?



g) Give an example of an undesirable model with high accuracy

h) For what is the Hoeffding bound useful?

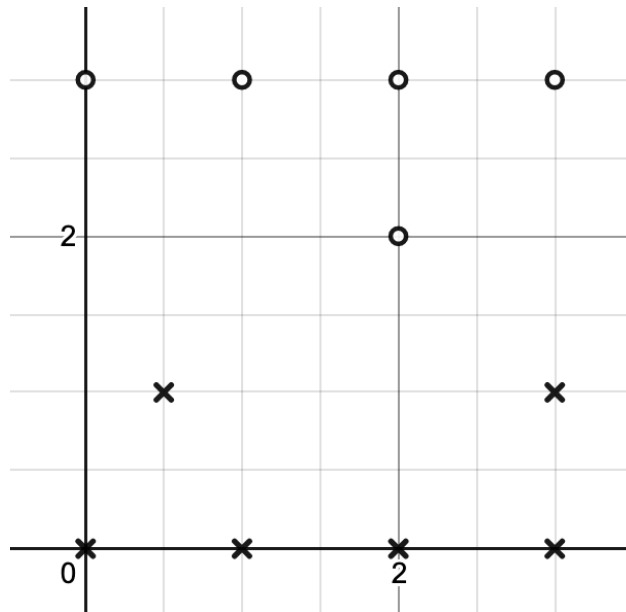
i) Why do we constraints do we put on the features in the Naive Bayes learner and why?

j) Why does a Random Forest usually have lower error than a decision tree? Why does it have lower variance?

### 3 Support Vector Machines

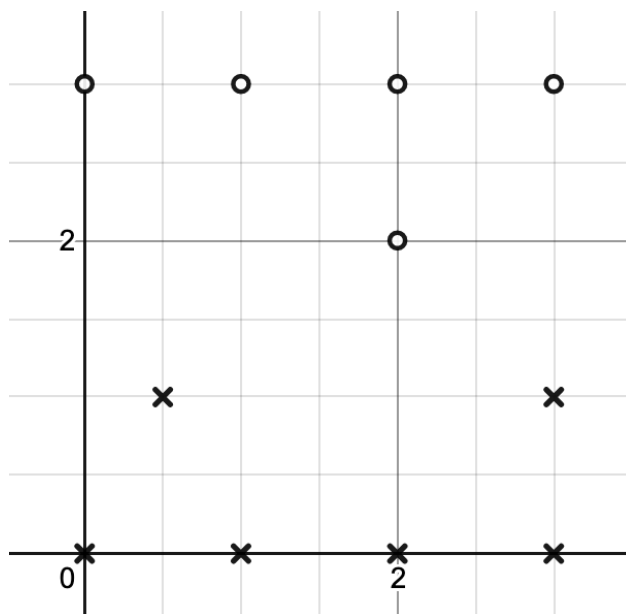
Answer the following questions related to Support Vector Machines. Let the "X" points be positive.

- a) Draw the hard-margin SVM on the dataset below. Draw the decision boundary as a solid line and the margin boundaries as dashed lines. Circle the support vectors



- b) What are  $\beta_0, \beta_1, \beta_2$  for the model from part a)?
- c) What is the hinge loss for the model from part a)?

- d) Consider an SVM with margin boundaries at  $x_2 = 1$  and  $x_2 = 2$ . Draw the decision boundary.

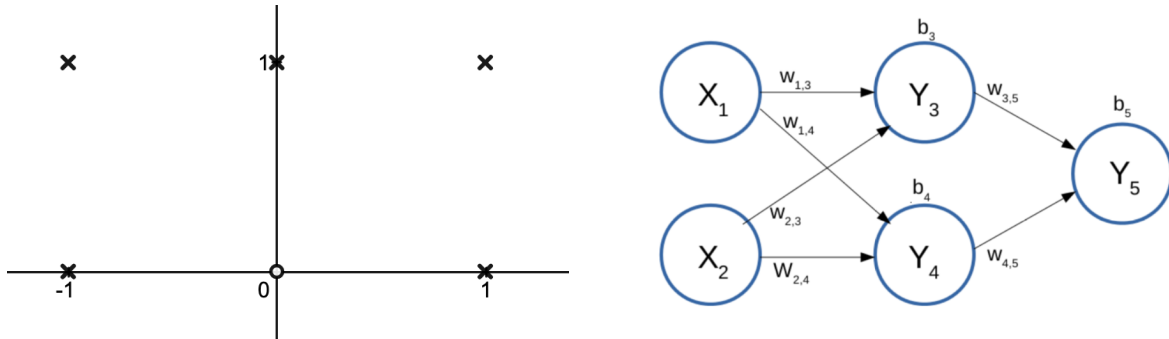


- e) What are  $\beta_0, \beta_1, \beta_2$  for the model from part d)?
- f) What the hinge loss of the model from part d)?
- g) For what values of  $c$  will a SVM prefer the model from part d) over the model from part a)

## 4 Neural Networks

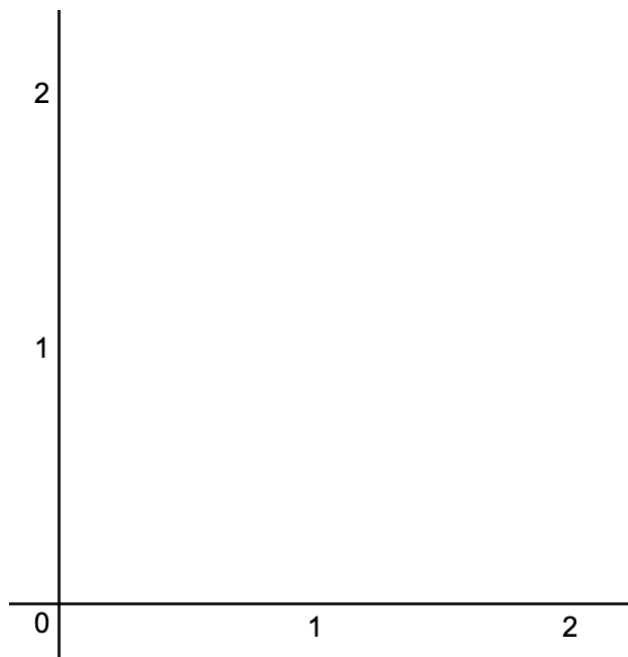
Consider the following data set and neural network model. Let the "O" points be positive. Use ReLU as your activation function

- a) Fit the following data with a neural network with all weights in  $\{-1, 0, 1\}$  such that the negative "X" points are all on the boundary  $y_5$ . Draw the boundaries of your internal nodes  $y_3$  and  $y_4$  on the graph. Be sure to indicate which direction is positive.



- b) Give the weights for your internal nodes  $y_3$  and  $y_4$ . All your weights should be in  $\{-1, 0, 1\}$

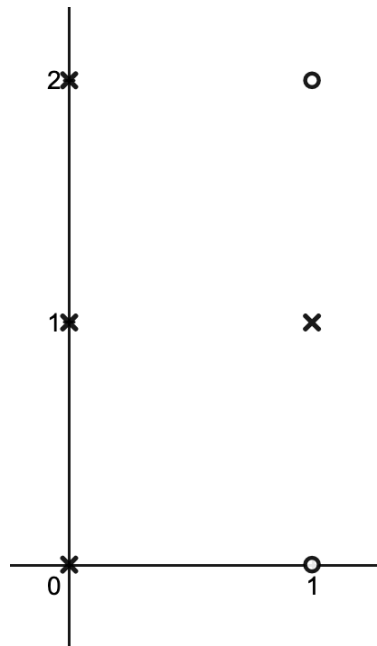
- c) Now draw the points in the transformed  $y_3, y_4$  space. Let  $y_3$  be the x-axis and  $y_4$  be the y-axis. Remember to use ReLU



- d) Draw your decision boundary  $y_5$  on the above graph such that all the negative "X" points lie on the boundary.
- e) Give your weights for  $y_5$ , that are all in  $\{-1, 0, 1\}$

## 5 Decision Trees

Consider the following dataset.



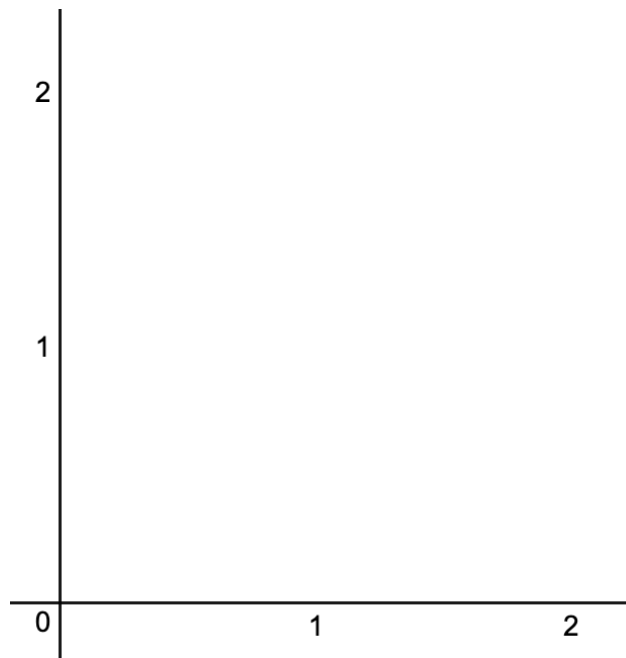
- a) Draw the best binary decision tree for the above data with one internal node

b) Draw the best binary decision tree for the data on the previous page with two internal nodes

c) Draw the binary decision tree which has zero training error.



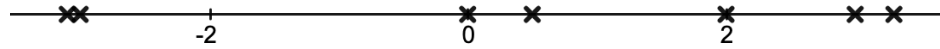
- d) Draw data points on the following graph where no horizontal or vertical separation of the data would cause a decrease in entropy.



- e) Describe an approach to post-pruning

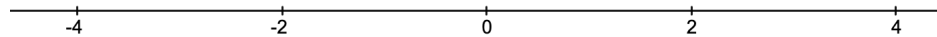
## 6 Unsupervised Learning

For this problem use the following linear dataset. Let the points be labeled A,B,C,D,E,F,G from left to right



- a) Draw the dendrogram for the following data set. Use closest points as your linkage method. Use agglomerative clustering

- b) Draw a dataset which has a different dendrogram for closest-point clustering and farthest-point clustering.



- c) Give both dendograms.

This page is left for you if you need extra space. If you use this space, please clearly mark on the problem page where we can find your solution.