Final

This is a preview of the published version of the quiz

Started: Jan 20 at 5:38pm

Quiz Instructions

This is the final exam for the course 422/622. You have 3 hours to complete the final.

Question 1 — 5 pts

Is there a situation where you would want to use a decision tree over a neural network? If so, give an example (be concrete). If not, explain why not.

Edit View Insert Format Tools Table

12pt Paragraph

If the data can be separated by axis-parallel lines, then there is no real need to use a neural network when a decision tree will give you an explainable result.

Basically, if you want your model to be explainable, you should use a decision tree instead of a neural network.

p 0 words

Question 2 — 5 pts

Name two advantages of a perceptron over an SVM. That is, why might you use a perceptron instead of an SVM?

Edit View Insert Format Tools Table

12pt Paragraph

1. The perceptron algorithm is simple and easy to implement.
2. The perceptron is quicker to train (even if you have to set a max number of epochs).

p 0 words

Question 3 5 pts

Learning to read is an example of reinforcement learning, imitation learning, or both? Explain.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 4

5 pts

A false positive pregnancy test is an example of bad precision or recall. Briefly explain.

Edit View Insert Format Tools Table

12pt Paragraph

False positive is an example of saying someone is pregnant when they are not, so it's bad precision, because you made a mistake.

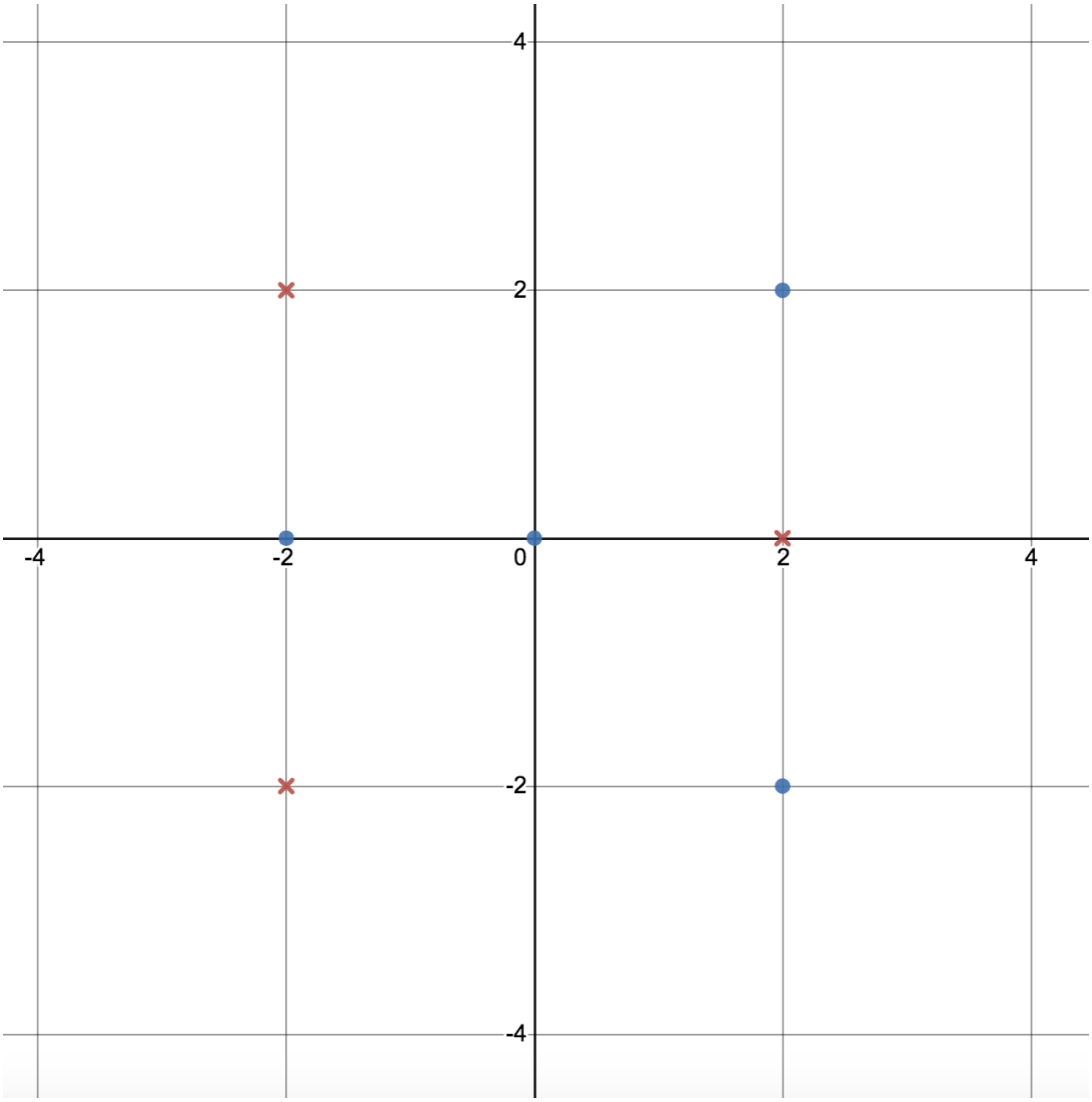
For it to be bad recall it would have to be a false negative.

p 0 words

Question 5

5 pts

For each of the following classifiers, state whether or not they will be able to correctly classify the following training. If the answer is no, explain why not. If the answer is yes, make sure to indicate the particular hyperparameters/setting under which the model would be able to classify the data.



a) Perceptron No. The data is not linearly separable.

b) Decision Tree Yes. No conflicting points. I think you could solve this with a depth 5 tree.

c) KNN Yes. KNN always achieves 100% accuracy on training data with K=1

d) K-Means No, if K=2. But if K=N, then it could effectively serve like KNN with K=1

Edit View Insert Format Tools Table

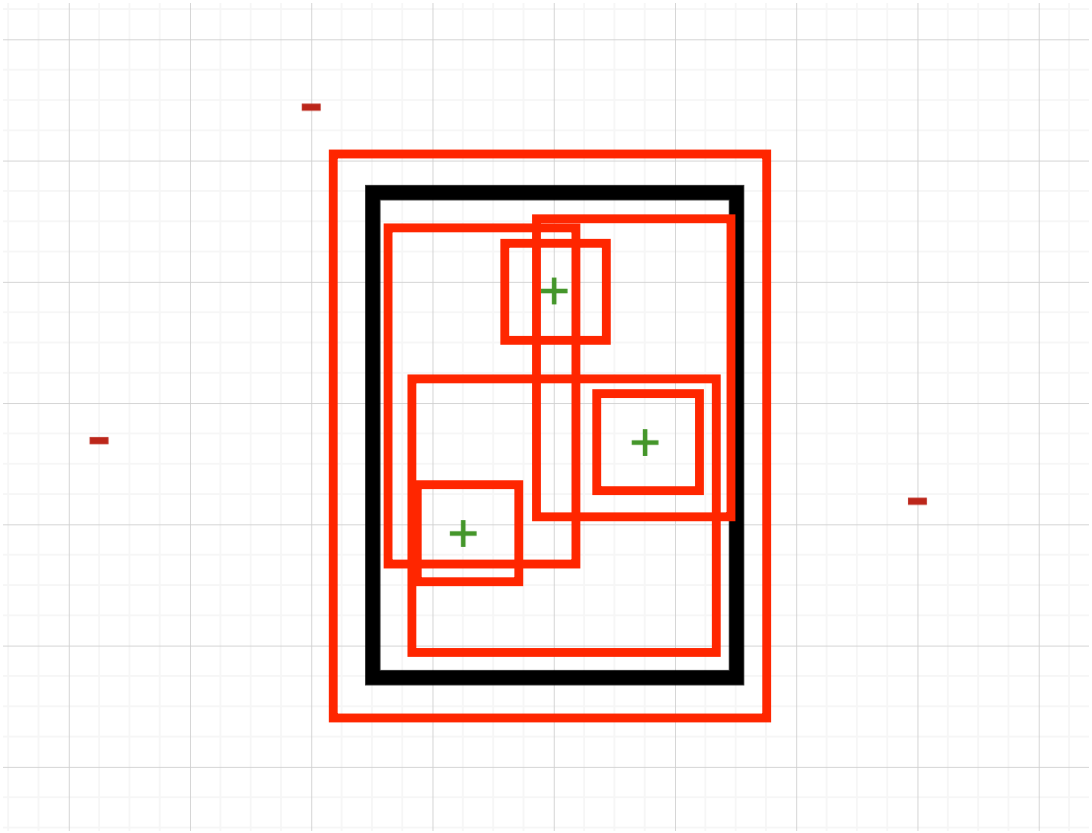
12pt Paragraph

p 0 words

Question 6

10 pts

Prove that the VC dimension of rectangle classifiers is at least 3. A rectangle classifier is simply a rectangle in 2D where the points inside of the rectangle are predicted as positive and the points outside of the rectangle are predicted as negative. See the image below as an example of a rectangle classifier. This is just an example so you can visualize the rectangle classifier.



Edit View Insert Format Tools Table

12pt Paragraph

The boxes above show how to separate each individual point, each pair of points and all the points. Allowing for a shattering of the data. So the VC dimension is at least 3.

p 0 words

Question 7

10 pts

Explain the similarities and differences between Imitation and Reinforcement Learning to someone outside of CS. That is, don't use technical jargon. Explain in plain English.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

Imitation learning is a very specific setting of reinforcement learning.

In imitation learning, you imitate an expert. For example, children use imitation learning to make expressions, copying what their parents do, turning their face into a smile or a frown. This is a type of learning through imitation.

In Reinforcement learning, you learn from interacting with your environment. Learning to walk is a good example of this. You try to walk and you fall, receiving feedback in the form of pain from the environment (you hit your butt or other body part on the ground causing pain). You may also receive other feedback from adults in the form of cheering when you are successful.

Imitation learning is simply a form of mimicry whereas reinforcement learning involves more complex feedback from the environment/other people.

p 0 words </>

Question 8

10 pts

Which of the following models can form a linear decision boundary? For each model, provide a brief explanation for your yes or no answer.

a) Decision Tree

b) K Means

c) K-NN

d) Neural Network

e) SVM

f) PCA

g) Kernels

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

a) Yes: Decision stumps are horizontal or vertical lines.
b) Yes. If K=2 you will get a linear boundary between the two clusters.
c) Yes, if there are two points and K=1. Or in some cases if K=N/2
d) No (assuming you have nonlinearity at the hidden nodes).
e) Yes. Assuming hard-margin SVM or soft-margin without Kernels.
f) No (no decision boundaries in PCA)
g) No. Kernels aren't a classifier.

p 0 words </>

Question 9

10 pts

Let's say I have a large dataset, with 20 million samples. Each sample is represented by a 15-dimensional feature vector consisting of real values. What would be the best ML model for this problem? What would be the worst? How would your answers change if the numbers were swapped (15 samples, with 20 million features each)? Explain.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

KNN would be terrible, because you'd have to hold all 20M samples in memory.
Neural network would probably be best or SVM. DT would be good, but not the best since there are real-valued features, which means you can split on a single feature many times.

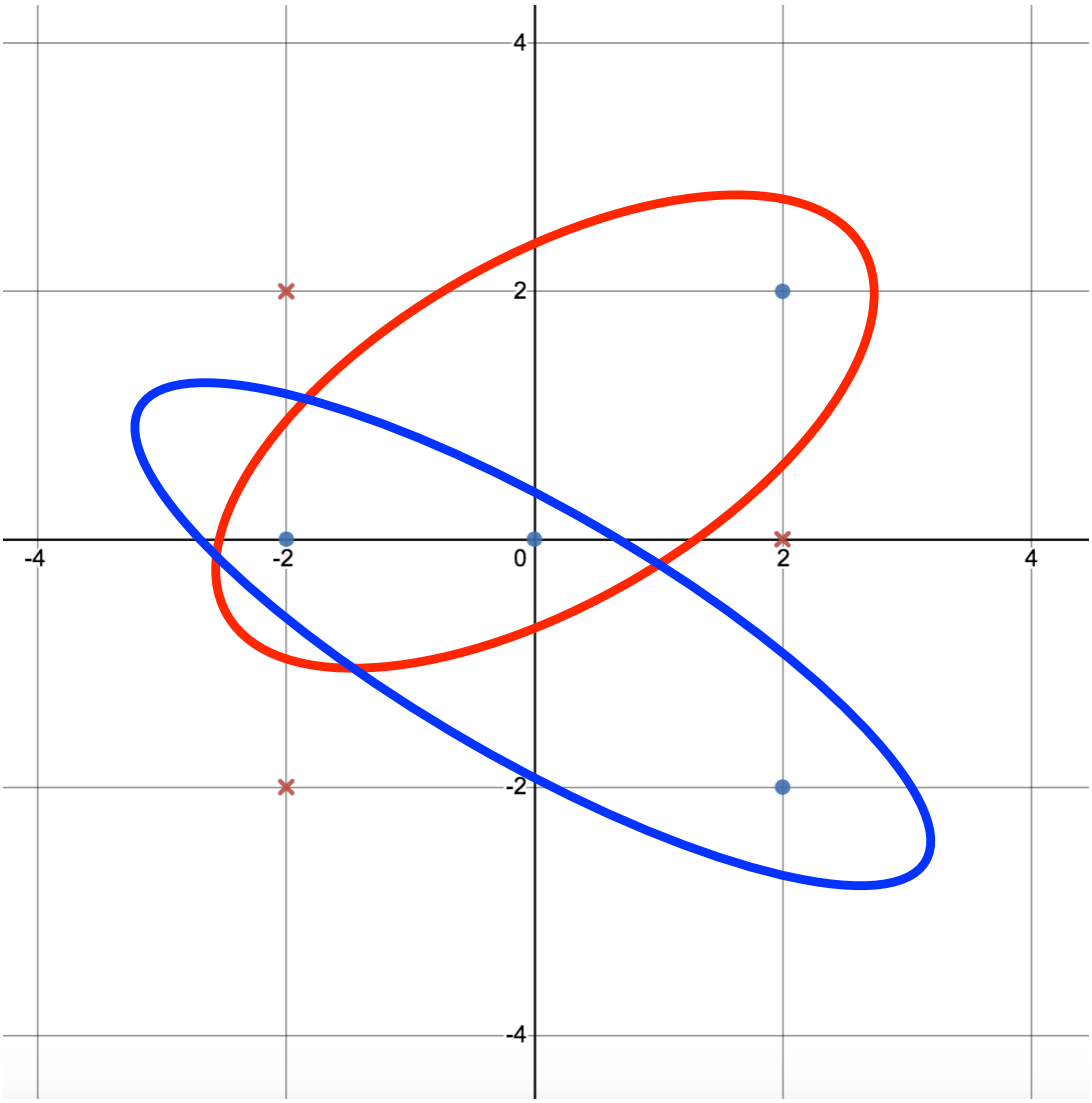
If we have 15 samples with 20M features, we still wouldn't want to use KNN, cause those are HUGE features. SVM and neural networks and Perceptron would be slow, since w grows with the # of features.
If you used Kernelized Perceptron, you would be better off cause alpha grows with the # of samples.
You would likely want to use a DT also since there are so few features your tree wouldn't be very deep.

p 0 words

Question 10

15 pts

Assume you have the following training data (red Xs are negative and blue circles are positive). We want to use Adaboost to train an ensemble classifier. Let's assume our classifiers are ellipses, where the points inside the ellipse are classified as positive and the points outside are classified as negative. Answer the following questions. You should be able to answer these questions without explicitly using the formulas in Adaboost. That is, if you're doing a bunch of math, you might be thinking about this the wrong way.



a) What classifier would Adaboost choose in the first iteration? Draw it. And explain in one sentence or less.

The red ellipse would be the first classifier, as it gets the minimum (1 sample) number of samples incorrect.

b) What would be the weights of each sample after the first iteration of Adaboost (after re-weighting)?

(2,-2) would have a weight of 1/2 and the remaining 6 samples would each have a weight of 1/12 so they sum to 1/2

c) What classifier would Adaboost choose in the second iteration? Draw it and explain in one sentence or less. The blue ellipse would be the second classifier. It would get the sample with 1/2 weight right and mis-classify one of the samples with 1/12 weight.

d) How would the final combination of the two classifiers classify the following point: (2,0)? Explain in one sentence or less.

The second classifier would have a lower error and therefore a higher alpha, So it would always be the classier that is used. So it would classify (2,0) as negative

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p

0 words

Question 11

20 pts

Given the following training data, there are some missing values (gray boxes). Each of the columns in the data represents a feature, except the first, which indicates a sample number. Assume you have labels for your prediction problem, they're just not in this figure. You can also assume there are more than 9 samples and this is just a small subset of the dataset. Answer the following questions.

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0		0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622		222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2		12.43
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15
8	0.21124	12.5	7.87	0.0		5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10

a) How could we use a K-NN classifier to fill in the missing feature values?

b) How could you use a neural network to fill in the missing feature values?

c) Is there another way to deal with the missing feature values, without removing the samples or trying to guess their values?

Edit View Insert Format Tools Table

12pt Paragraph

a) You can train a KNN classifier and use the sample's nearest neighbors to fill in the value. So remove that column all together and find the K-NN of the sample and take the average of those nearest neighbors to fill in the missing value.

b) You can train a neural network to predict the missing feature value.

c) You can simply replace them with some null value (0).

p

0 words

Not saved Submit Quiz