

CS 412

APR 7TH – EXAM REVIEW

Review

Midterm, Thursday April 9th, 12-8pm CDT

- Previous exams posted
- Video solutions up tonight

4 questions (100 points)

→ Short Answer

SVM

NN

Boosting

72hrs

Post on Piazza.
Submit on gradescope

large work by
hand
One will
have
a small
programming
portion
Likely to be mostly
very simple

Short Answer

8 questions

- 5 points per question
- Definitions and qualitative answers

Open-book
open-note

Why might we
prefer one model
to another
Defend your answer!
bias v. variance

Short Answers

SP14Final: Q1.4: If the non-support examples of an SVM are removed from the training set, will the SVM produce the same output model?

Yes.

All removed
points are
not Support Vectors

SVM

Support vectors:
any point on
the margin boundary
or any point w/
high loss

Short Answers

anytime
we pick the
best, we
introduce
bias

SP14Final: Q2.1: Given a classification method with some model parameters that we need to choose, how should the performance on unseen data be estimated using an available dataset? Describe specifically which data should be used

testing vs. training
See what the data

```
graph TD; A[See what the data] --> B[test]; A --> C[train]; C --> D[Cross]
```

validation
before & after
use D^* on test data
for our parameters
error
unbiased
estimator

Short Answers

Give two tuning parameters for each of the following model types:

- Polynomial kernel SVM

degree, C

- Neural Network

nodes, layers

- Nearest Neighbor

k , weights

- Adaboost

Documentation
lookup
Scikit-learn

Short Answers

In general, bagging is an ensemble technique used to reduce Variance
and boosting reduces bias

why?
Independence

$$T = 1/C$$

Logistic regression
 $1 - regularization$
parameter

C-
(one)

Short Answers

Spr15Final: Q1.3: Will a logistic regression model with L1 regularization will have lower training error than the same model without regularization? Explain.

Yes
w/o regularization
L1 finds the max
w/ minimum training error
Other models
cannot be
better

min

L1: $\sum w_i$

L2: $\sum w_i^2$

error + regularization

maybe
case that
L1 error
is lower
if L1 is
overfit

Short Answers

Fall15Midterm: Q2.1: Describe overfitting and give a situation in which it is likely to arise

underfitting - or model is insufficiently complex

overfitting ← fitting to the noise
correct w/ many parameters
via cross val

Short Answers

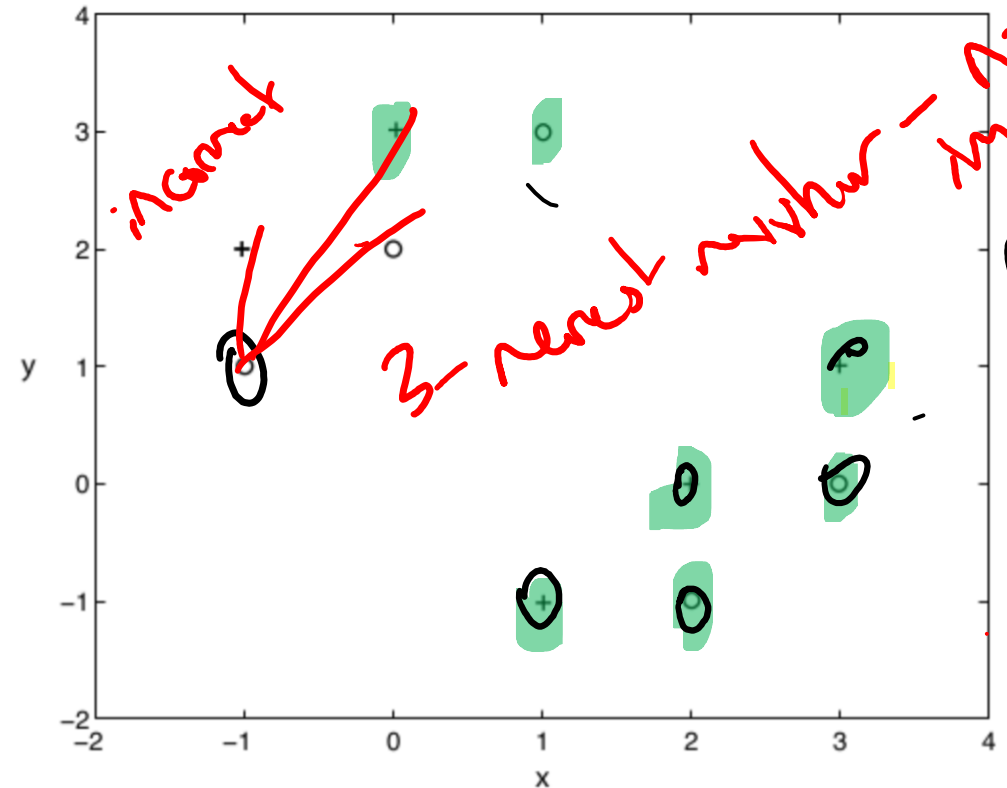
Spring13Midterm: Q2.: What is the leave-one-out-cross-validation error for the following data set for the 1-Nearest neighbor model?

Which value of k has less error, explain your answer

1/2 fold: 1 error
9/10 folds: 16/10

1-fold

LOOCV - why for KNN



Boosting

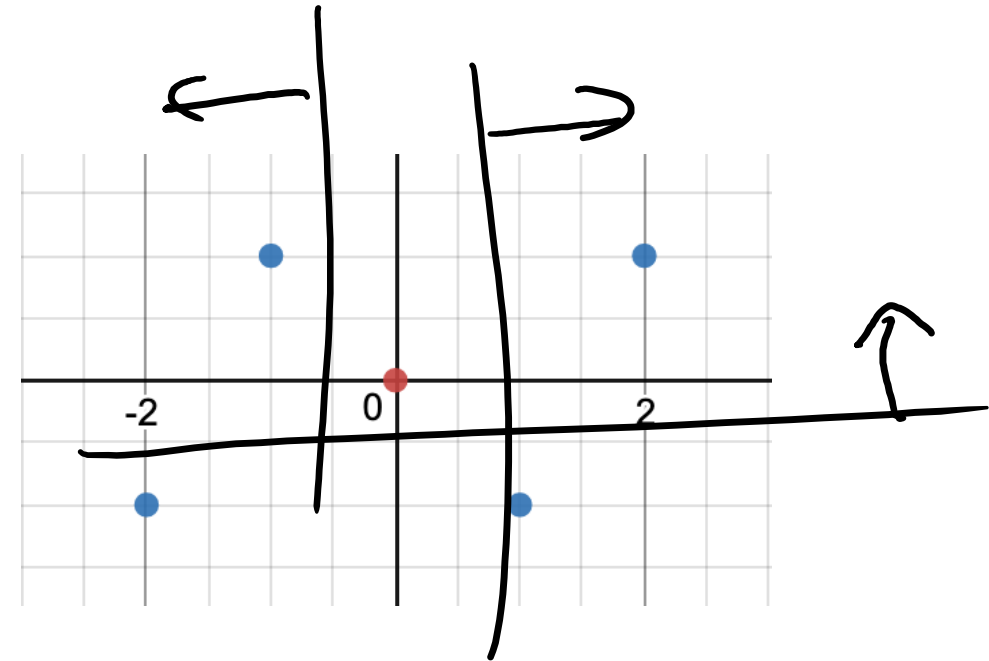
Perform 3 iterations of the boosting algorithm on the following dataset. Double the error at each point and weight each model with $\alpha_i = (1-\epsilon)/\epsilon$

Programming

→ sci-kit learn packages

→ plot simple plot

↳ exactly a plot I've had
you to before



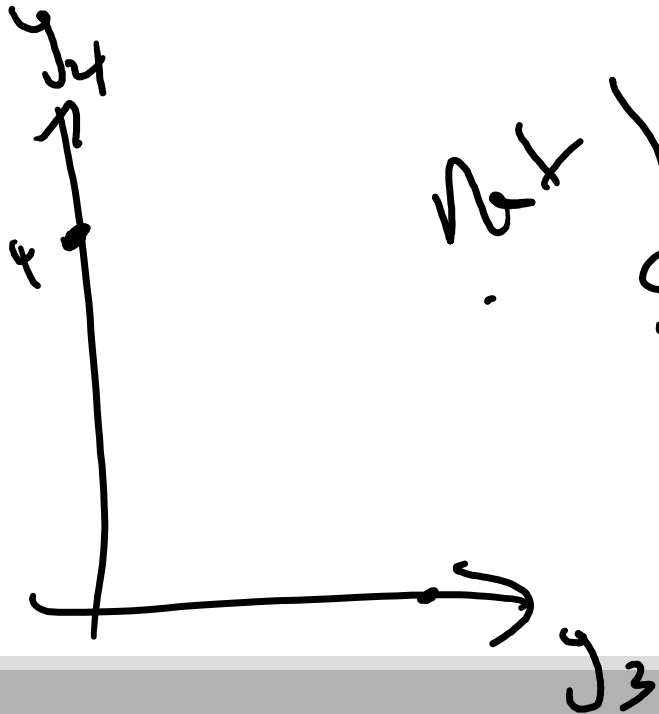
Discussion!

Neural networks

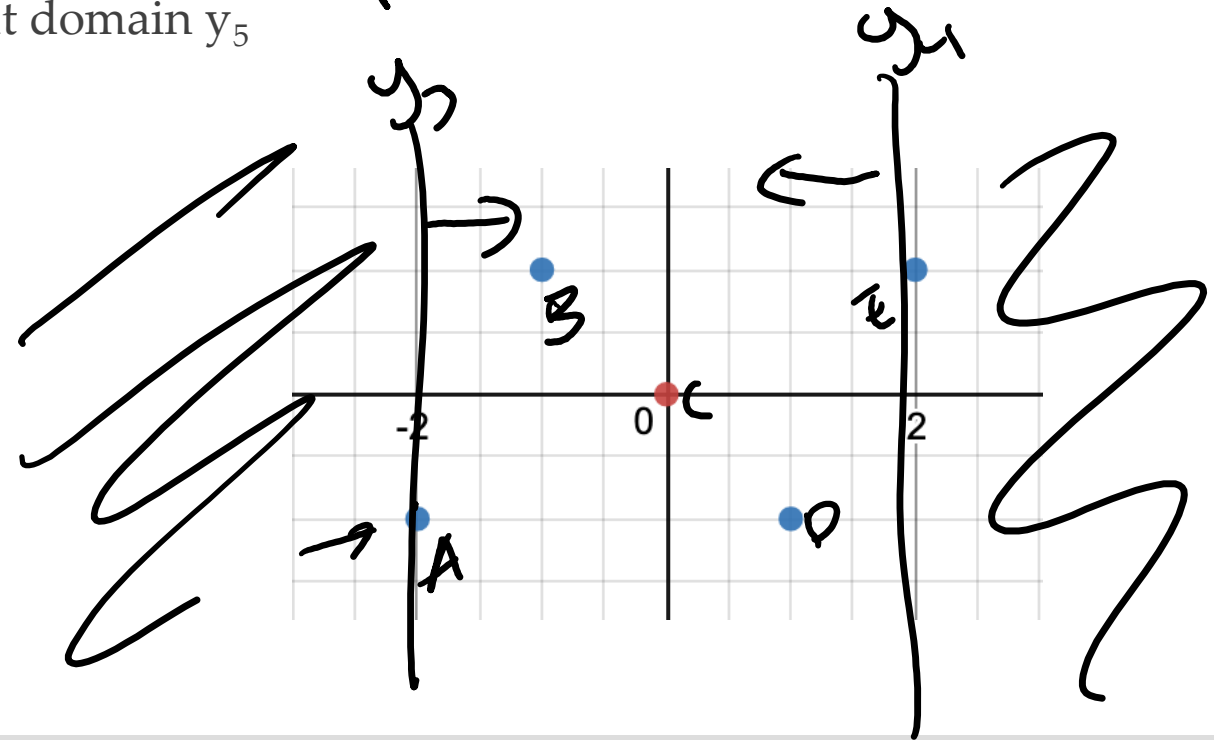


→ Fit the following data with a neural network with one hidden layer of 2 nodes (y_3, y_4) such that the positive point (red) lies on the decision boundary. Use ReLU as your activation function and let all the weights be in $\{-1, 0, 1\}$

Show your y_3, y_4 plot and your final decider in that domain y_5



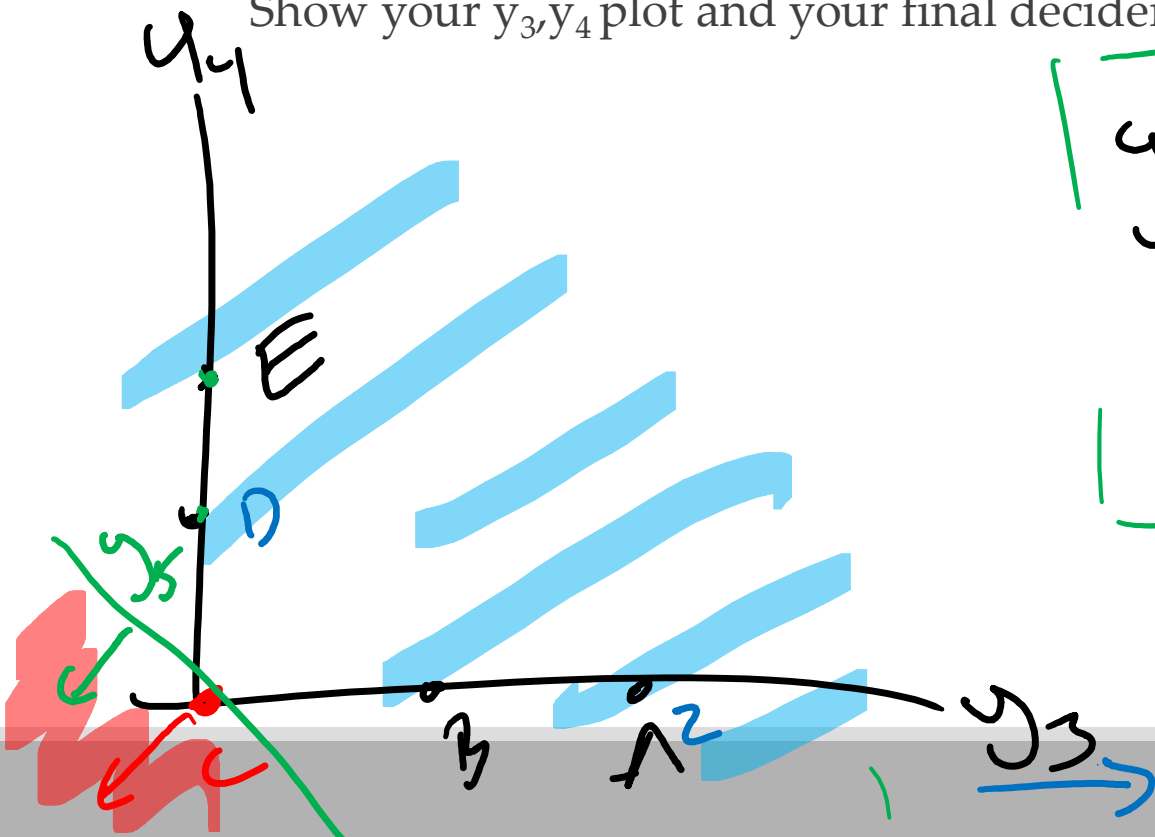
Not linearly separable



Neural networks

Fit the following data with a neural network with one hidden layer of 2 nodes (y_3, y_4) such that the positive point (red) lies on the decision boundary. Use ReLU as your activation function and let all the weights be in $\{-1, 0, 1\}$

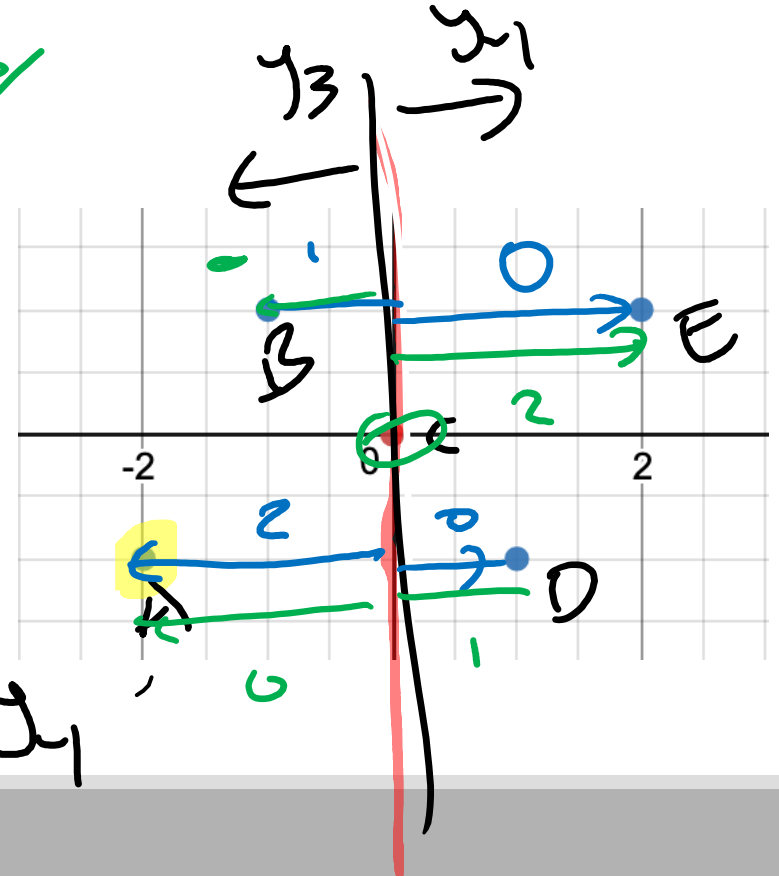
Show your y_3, y_4 plot and your final decider in that domain y_4 .



$$y_3 = x_1$$

$$y_t = x_t$$

$$y_5 = -y_3 - y_1$$



Support vector: point on margin OR
point w/ highest

$$y = w_0 + w_1 x_1 + w_2 x_2$$

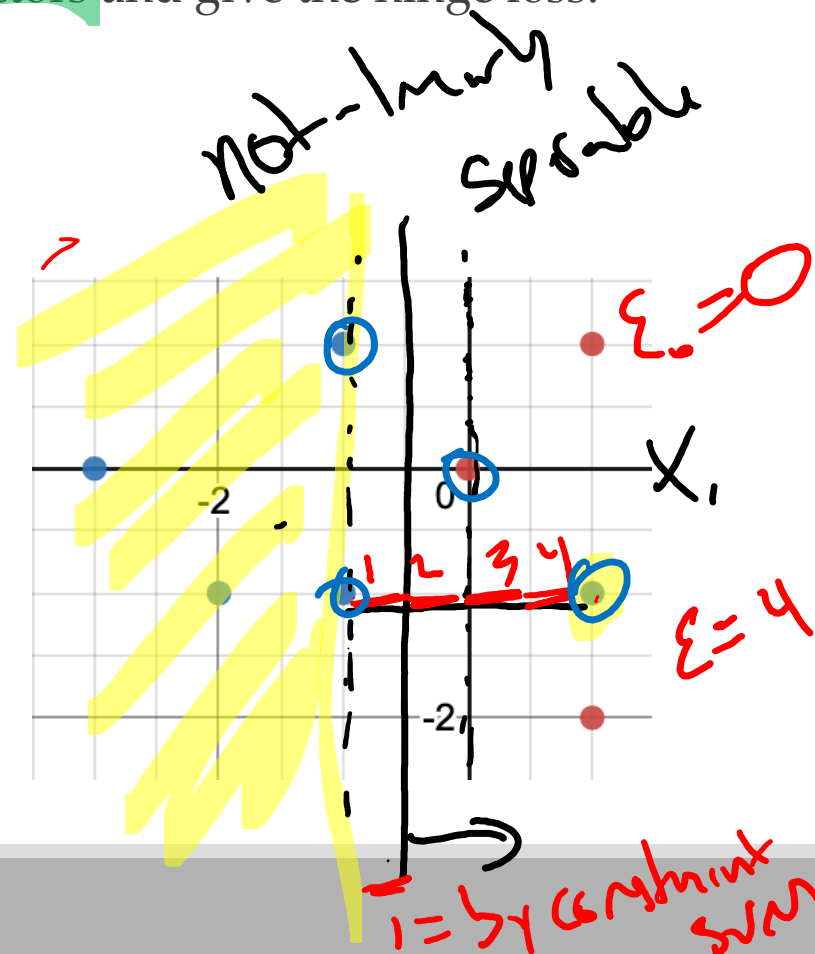
$$u = 2$$

ψ_0

$$x_1 = -1/2$$

$$0 = \omega_0 + \frac{1}{2} \omega_1$$

$$1 = w_0 + w_1$$



\rightarrow
1 = by constraint sum

SVM

$$|S_v| = 6$$

C70

Consider the following data set. Draw the decision boundary and margin boundaries for an SVM with minimal hinge loss. Indicate which points are your support vectors and give the hinge loss. Let the red points be positive

Compare this to the SVM classified as $y = x_1$

What is the decision boundary/region boundary for this SVM?

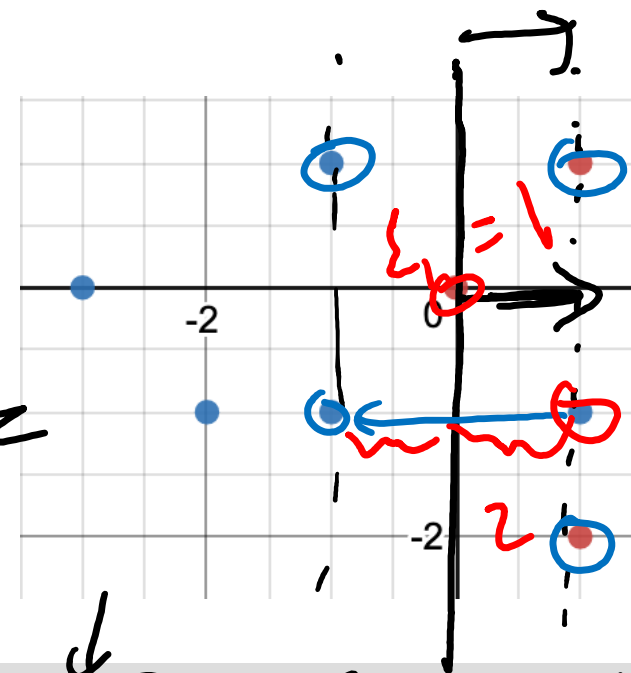
What are its support vectors?

What is its hinge loss?

For what values of c might we prefer this model to our original?

$$\frac{\|w\|^2}{2} + C \sum \epsilon = \frac{\|w\|^2}{2} + C \sum \epsilon$$

$$\frac{2^2 + 1^2}{2} + C \cdot 4 = \frac{5}{2} + 4C = \frac{1}{2} + 3C$$



select

shis

speed
qua
shock
resist

Good luck

Open-book, open-note

← all work should be your own

Previous exam

Video solution up by tonight

[HW2+3 regrades]

Project

Work up Dec on Friday

↳ team members (up to 3)

↳ rough project idea

12-8

expectn \leq 2hrs

HW4 out
Thursday

Dec in 2 weeks
Ensemble