CS 412

APRIL 30TH – ETHICS IN MACHINE LEARNING

HTTPS://HAL.ARCHIVES-OUVERTES.FR/HAL-01724307/DOCUMENT

Link

HW5 due Tonight
OH 5-7pm

Final
next Wednesday
12:01am - 11:59pm

Project Due next Friday
No Lab Days

Monday for
project meetings

Role of ML in Society

Computer-aided decision making vs. automated decision making

- *Which is better/worse?*

More and more jobs in ML and data science

- *Level of "expertise" required has gone up and down*

Impacts from automation impact on non-technological fields

- *Job migration*

↳ big loss of domain info

↳ we now have non-CS doing ML

Personal Experience

Disclaimer: not all ML has high stakes ethical concerns, however you should **always** be considering it

State of Washington

- Foster care placement
- CPS interventions
- Abuse prevention

Collaboration with Federal data scientists

Most bad ML is not malicious, but incompetent

- This is never an excuse.

Seven Main Ethical Concerns of ML

1. Data bias
2. Autonomy
3. Explainability
4. Decision-making
5. Consent
6. Responsibility
7. Privacy

Data Bias

Data collection is a nebulous process

Biases can be both overt and hidden

- Decision making can be based on a process of unethical information
- This data can also be encoded in other "safe" variables

Data fairness

- Beyond just providing the "best" model, which model provides the most fair result
- Large field of new research, especially around public ML models

Example: NYPD COMPSTAT + Stop and Frisk

Correcting Data Bias

1. Maintain high quality data collection
    1. Longitudinal study
2. Ensure that data mirrors the diversity of the use population
3. Eliminate or confound variables with social discrimination
4. Record data collection and use practices

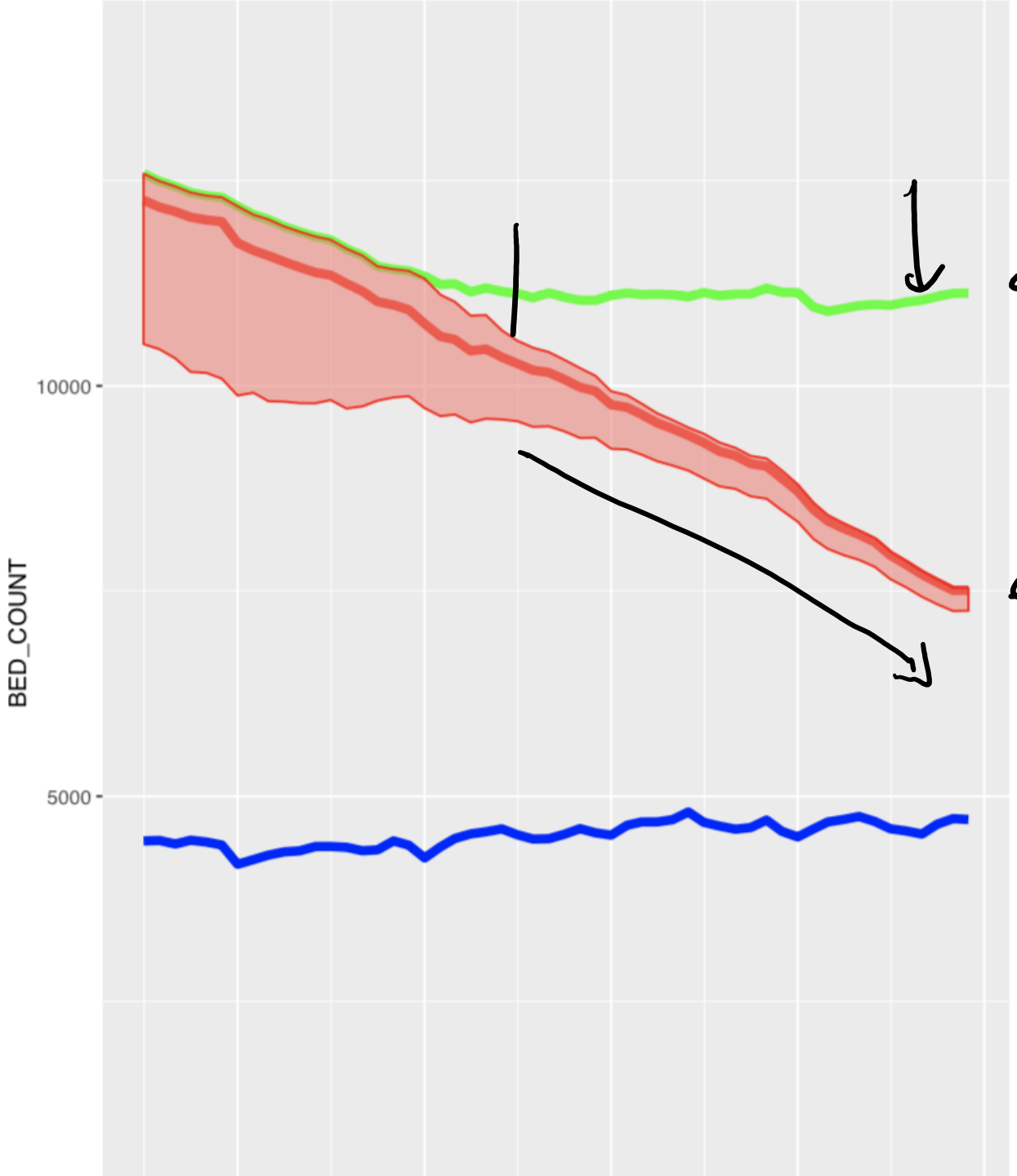
Autonomy

Computer-aided decision making
- ML helps to present relevant information to a human user who interprets the data to make a decision
- Safer, but less "desirable"

Automated decision making
- Different ethical concerns
- Explainability + responsibility

Misleading information
- Data presented to decision makers needs to accurately represent the data

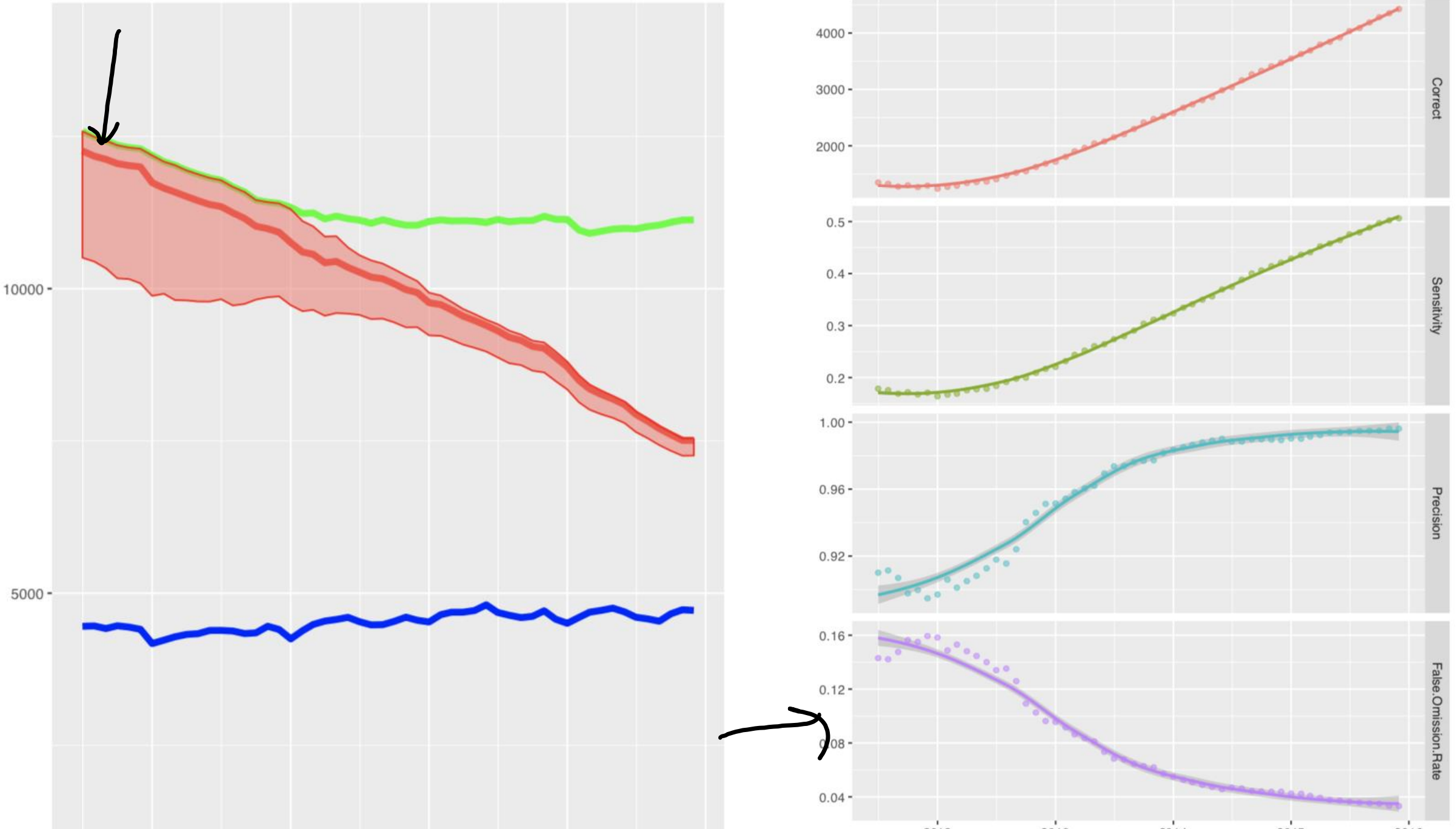
Example: Boeing 737 MAX



available beds

← likely beds → predictor

used beds


10000
5000
4000
3000
2000
Correct
0.5
0.4
0.3
0.2
Sensitivity
1.00
0.96
0.92
Precision
0.16
0.12
0.08
0.04
False.Omission.Rate

Explainability

Different disciplines have different standards for explainability

To what extend does a model need to be transparent?

Definition:

- Sufficiently documented
- Comprehensible to **all** users
- Formally verified

Especially concerning for unsupervised learning

Can we trust machines to automate the explanation process?

Can I see how a decision is made?

no labels + little accuracy feedback

big area of research introduce autonomy concerns

Decision making

To what extent does human moderation impact decision making?

how many people impact the complete work

Is there open communication to people who will be affected by the decision making process?

Humans take an important role in understanding the decision process and also in ensuring that the process is open and transparent

How you made your model is just as important as the data itself and should be freely available

Example: Ohio voter roll

Consent — Data collection level - IRB

Data given in an online environment often has soft ethical notation of consent

We think it's reasonable for Amazon to recommend a purchase, and it's based on the same level of online data collection that Facebook uses to categorize users into "marketable" divisions

Legal definitions here are well below ethical expectations

Even with medical data, especially for sparse data points, should we allow a computer to cross-apply information from one patient to another?

Additionally, do users consent to have decisions made by a computer that impact them?

Example: Facebook, Facebook, Facebook!

Responsibility

A computer cannot take responsibility for its decision making process

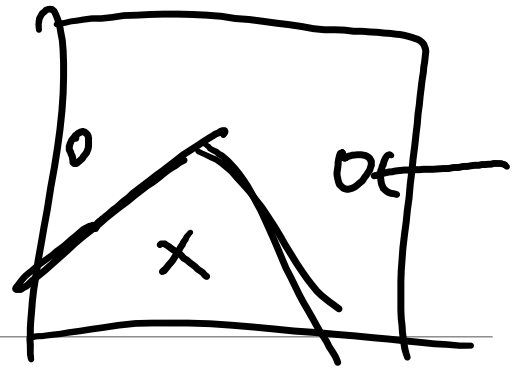
Sometimes, this is on purpose, but it is usually coincidental

As a potential designer of an ML system, you should be prepared to take responsibility for the decisions your model makes

Did you build the ML model with the intent of getting a certain result? ← documentation

Example: Bank loans/housing → bias

Privacy



To what extent does your model protect the data/identity of the data points used within it

ML models are getting access to more and more information that is more and more private

Huge financial incentives to ignore ethics and privacy concerns

How do you ensure that the data provides the best explanation possible, yet still protects individuals who provide the information?

Example: Census bureau

Prescriptions

Are there any procedural ways that we can mitigate ethical concerns?

Prescriptions

Are there any procedural ways that we can mitigate ethical concerns?

- Transparency and Documentation
  - Make sure the model is interpretable and that your documentation is readable by **any** user

Prescriptions

Are there any procedural ways that we can mitigate ethical concerns?

- Transparency and Documentation
  - Make sure the model is interpretable and that your documentation is readable by **any** user
- Communication
  - Machine learning never occurs in a vacuum. Discuss with your coworkers any concerns and problems that you have
  - ML is no place to be self-conscious—you need to be open when something is outside of your ability

Prescriptions

Are there any procedural ways that we can mitigate ethical concerns?

- Transparency and Documentation
  - Make sure the model is interpretable and that your documentation is readable by **any** user
- Communication
  - Machine learning never occurs in a vacuum. Discuss with your coworkers any concerns and problems that you have
  - ML is no place to be self-conscious—you need to be open when something is outside of your ability
- Social awareness
  - Even "low impact" ML can have real world consequences
  - This doesn't mean that we shouldn't do ML, but we should be prepared for our models to be discussed publicly

Prescriptions

Are there any procedural ways that we can mitigate ethical concerns?

- Transparency and Documentation
  - Make sure the model is interpretable and that your documentation is readable by **any** user
- Communication
  - Machine learning never occurs in a vacuum. Discuss with your coworkers any concerns and problems that you have
  - ML is no place to be self-conscious—you need to be open when something is outside of your ability
- Social awareness
  - Even "low impact" ML can have real world consequences
  - This doesn't mean that we shouldn't do ML, but we should be prepared for our models to be discussed publicly
- Human decision making
  - To what extent is it reasonable for a human to guide the process?
  - As an ML user, are you getting all the data you can get from your model to make an informed decision?
  - If the model is continuously updating, you need continuous oversight. Even if it is fixed, uses might change

Questions to ask

Who stands to benefit from this model? What are their incentives?

What answer did we expect the model to have? Did we allow ourselves to allow models that deviate from our expectation?

To what extent have you involuntarily impacted the output with your process? Which data points are changing?

Have you accounted for randomness? Is the random error equally distributed?

Are users aware that ML is affecting them? Is there any communication?

What open-source code did you use? Is it reliable? Is it safe given privacy concerns?

Have you as the designer adequately communicated the model's flaws?

Ethics Conclusion

Prediction and ethics are inseparable
- No process is errorless

Know the difference between negligence and malice
- Everything you document should be evidence you were neither
- ML saturation ←

Self-regulation
- How do you look objectively at your own work?

↳ Corollary question

Final Exam

Wednesday May 6th 12:01 am – 11:59 pm. Open-book, open-note, open computer

1. True/False w/ explanations (5 questions) 10 points
2. Short Answer (8 questions) (30 points) ← Cumulative
3. Decision Trees (10 points)
4. Clustering (10 points)
5. Client specification (30 points) ↙
6. Faulty Experiment/Design (10 points)

Final Exam – Mock exam on piazza

5 True/False questions
- 1 point for correct answer
- 2 points for correct explanation

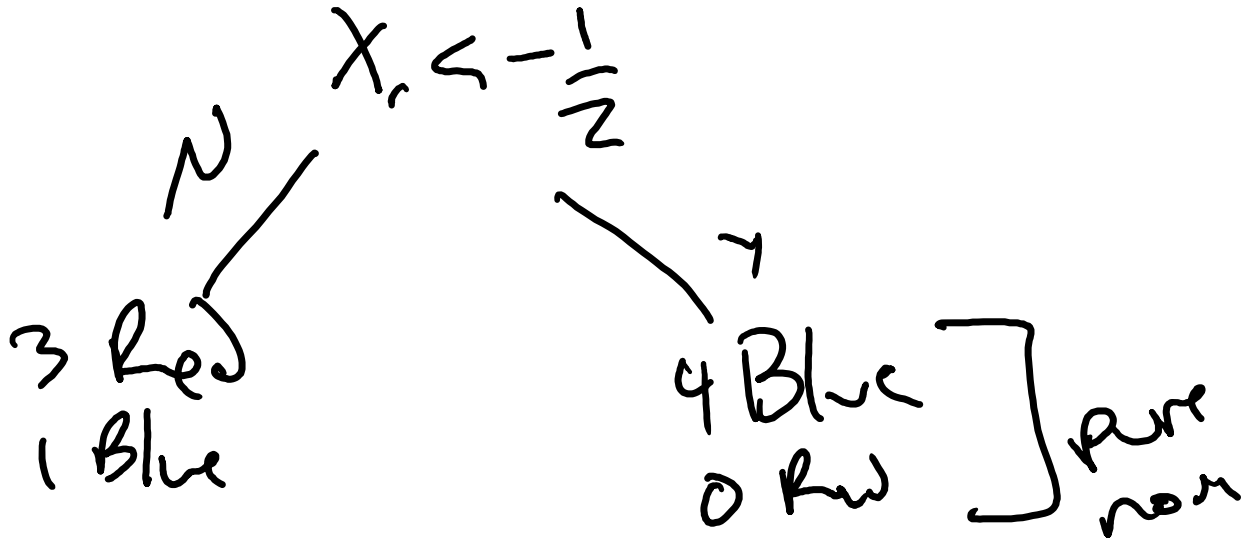
Short Answer
- 8 questions
- Ethics
- Concentration bounds

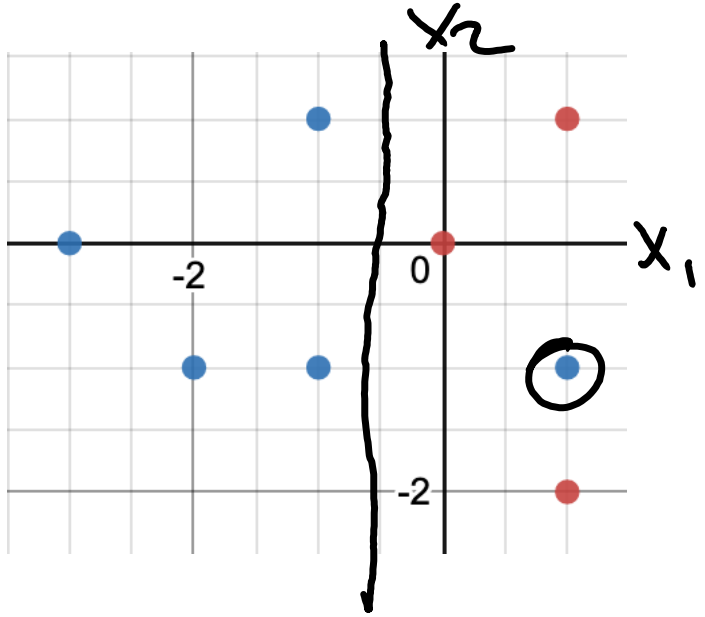
Markov
Chebyshev
Hoeffding

Decision Trees

Produce the decision tree of depth one for the following data points and give its error

or maybe entropy

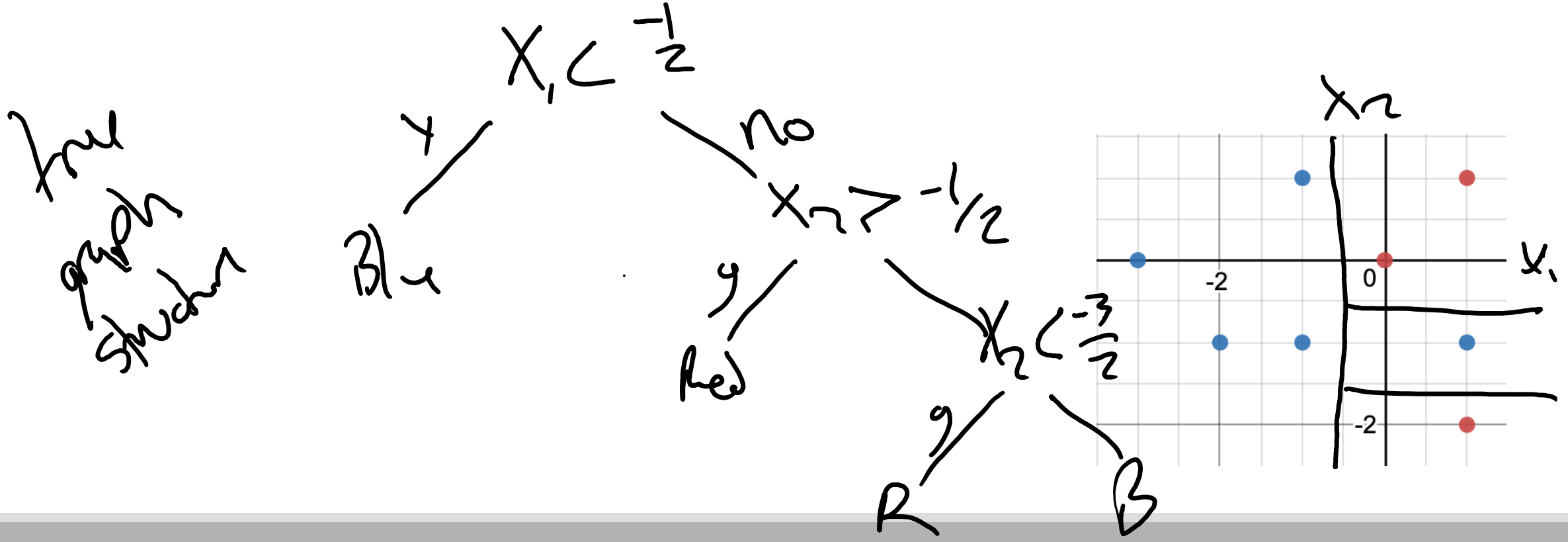




Decision Trees

Produce a decision tree which has zero training error.

(it does not need to be the decision tree produced by the entropy approach)

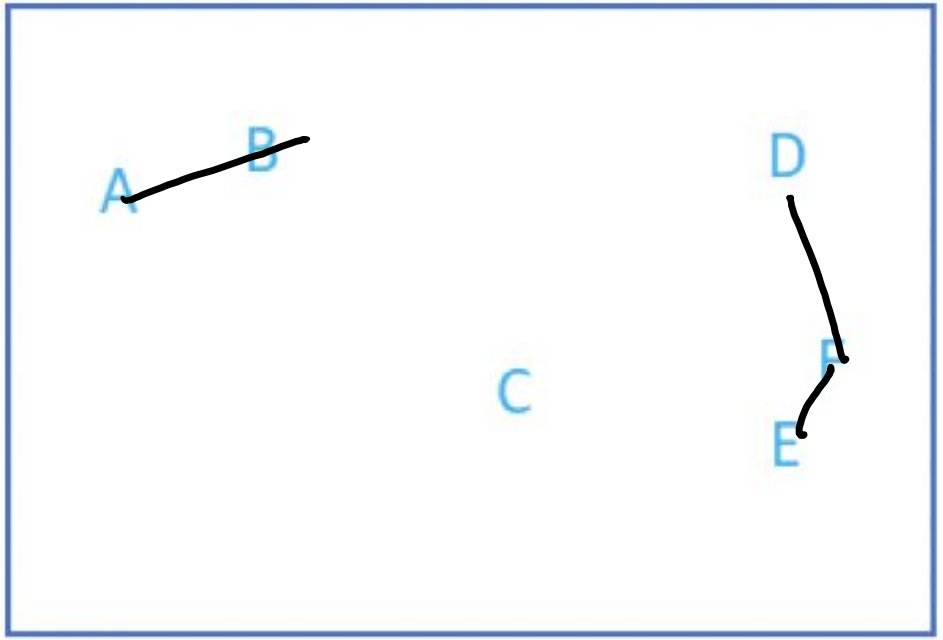


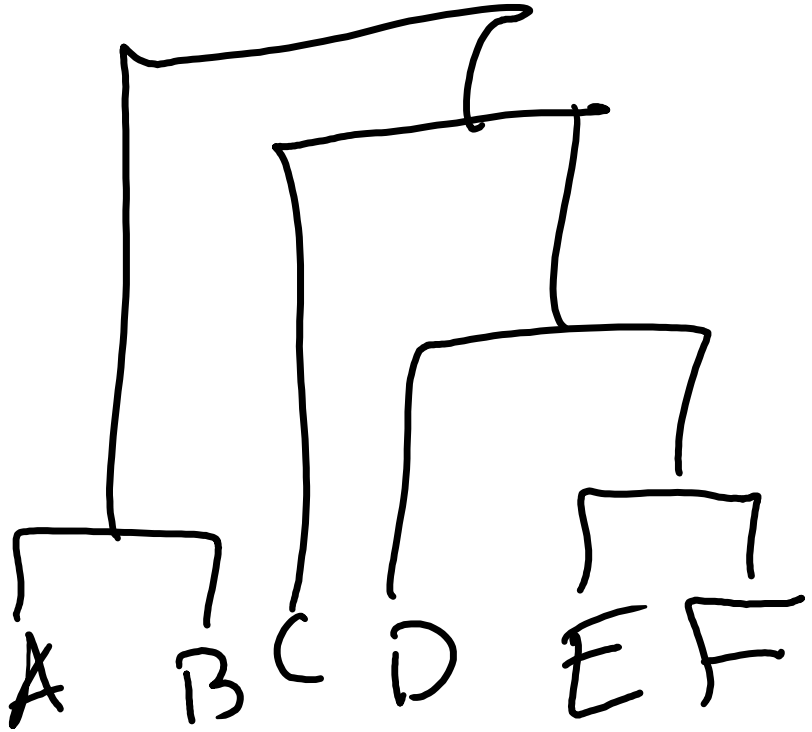
Clustering

agglomerative

Give the dendogram for the following data points. Use closest-points as your linking scheme

- Give a best estimate if points seem close (C,D)

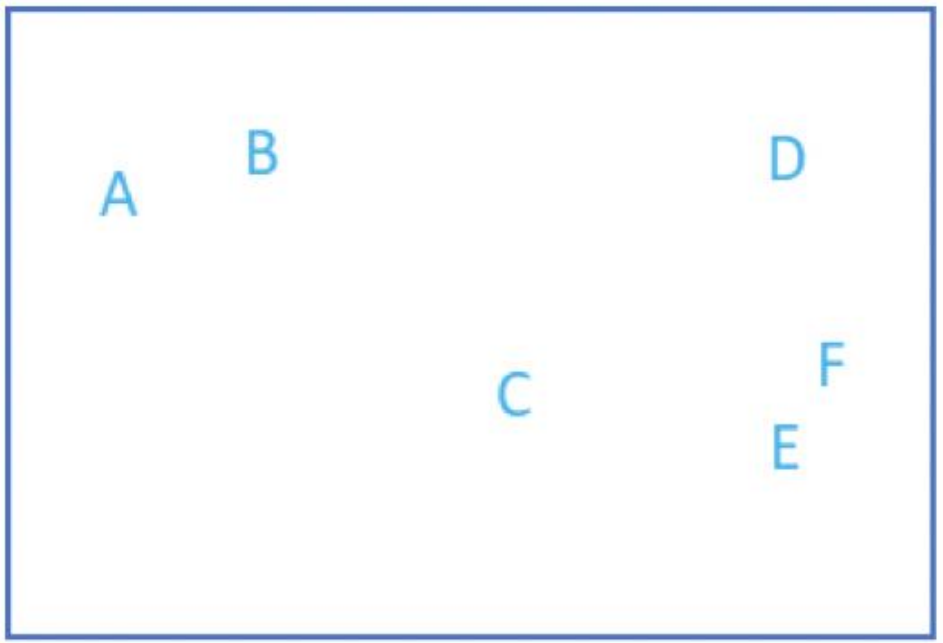


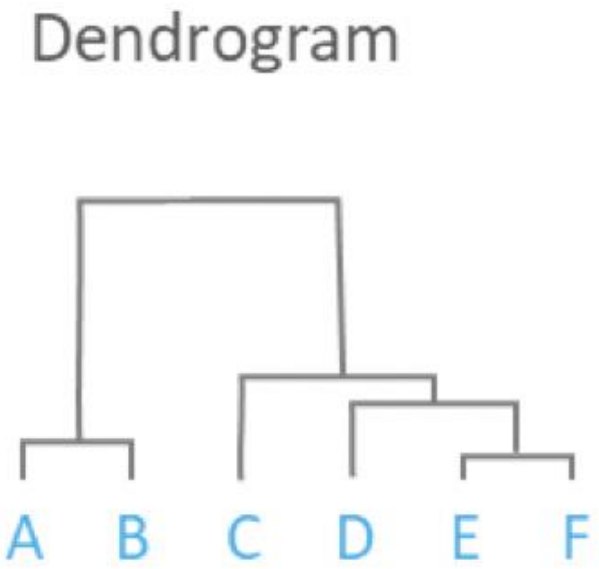


Clustering

I may use 1D data

Give the dendogram for the following data points. Use closest-points as your linking scheme

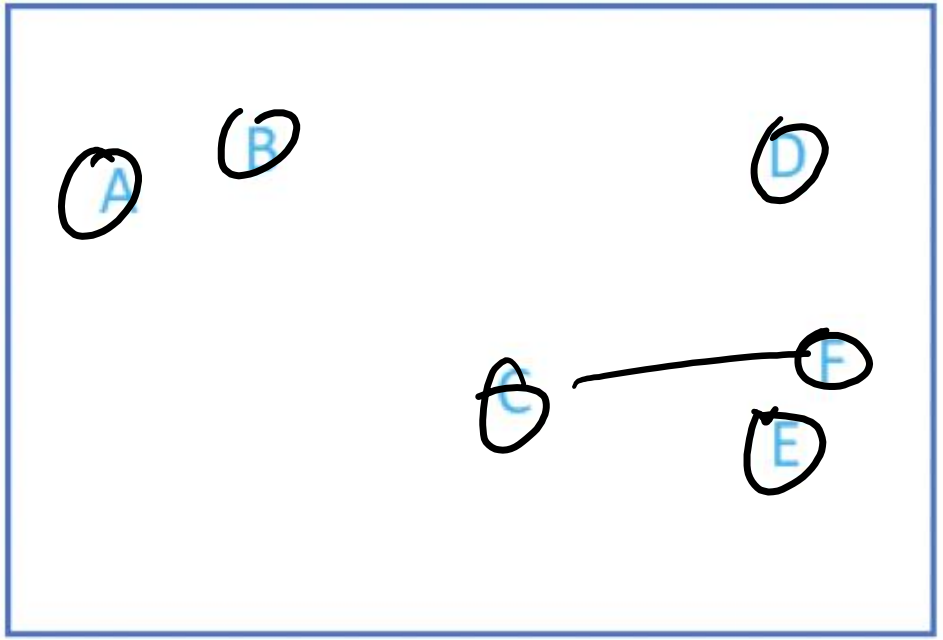


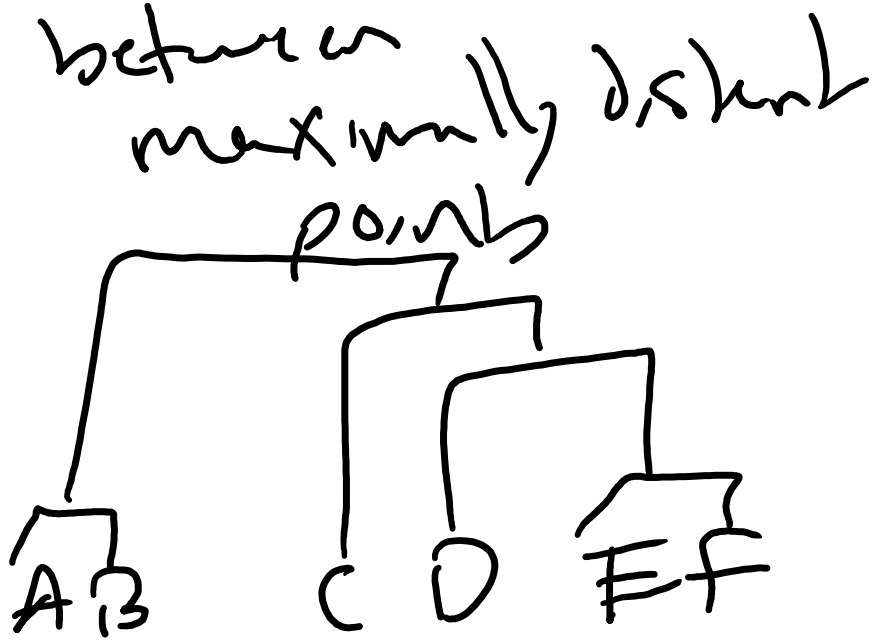


Clustering

Give the dendogram for the following data points. Use complete linkage as your linking scheme

- Give a best estimate if points seem close





Client Specification

For this problem, you will be given a new data set and some client specifications

You will need to:

- Find the model that best fits the client's needs
- Justify your decisions
- Give bounded error

only consider HW 4 models

Example specifications

- Cost asymmetry
- Need for explanatory power
- Small data set
- Lots of non-relevant features

ethics

~2 pages, include figures

Faulty Experiment

You will be given a short two page memo with figures defending a ML process

- Find all errors
  - Ethical
  - Statistical
  - Logical

- You will not need to redesign the experiment, but you will need to explain why you believe something is in error.

Finale

Please fill out course evaluations!

Also, regrade requests for all currently graded assignments due tomorrow

- Please check blackboard to make sure the grades look accurate

Meetings on Monday - sign up on the form

Final next Wednesday

Project Due Friday

Long Labs-yes the next

Final percentages posted to blackboard by Tuesday morning May 12

- Last chance for grade corrections before grades go to the registrar on Wednesday

thursday > QS DT

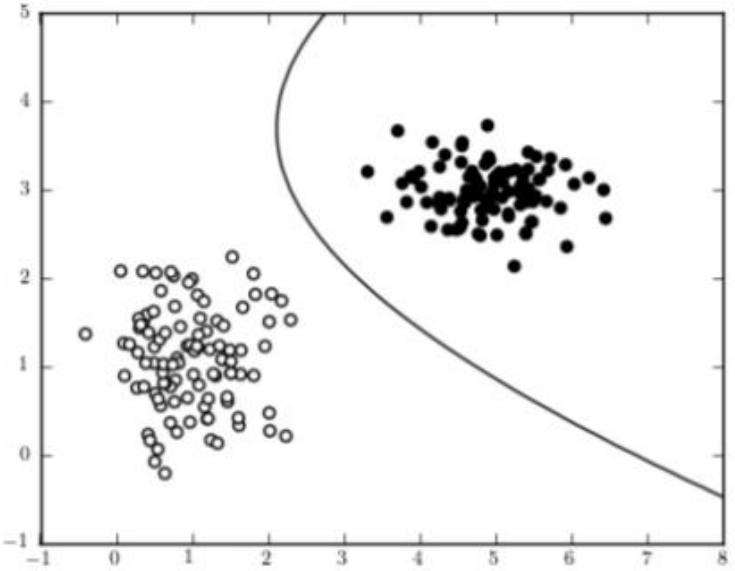
QS Clusters

Post solutions

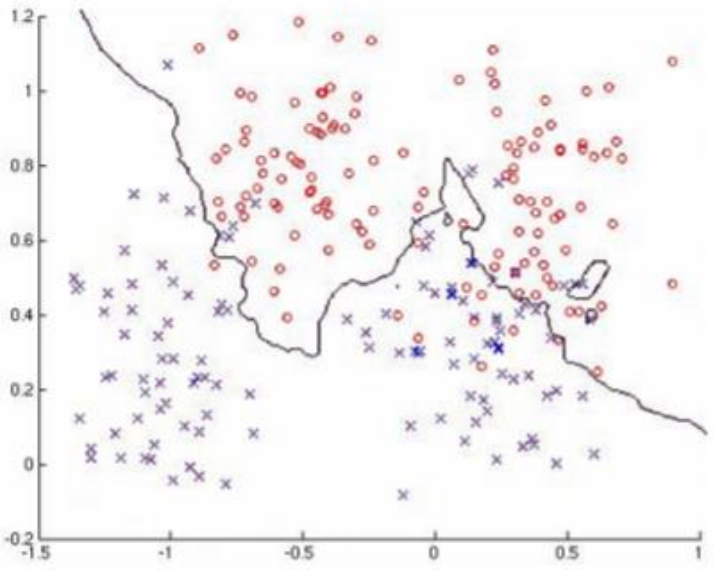
+ to MC + SA

Q2.4: (12 points) Consider the folowing predictors: decision tree, k-nearest neighbor, Gaussian naïve Bayes, logistic regression that is linear in the inputs, and linear support vector machines. You will need to explain which predictor(s) could produce a particular decision boundary and the parameters of the predictor that would do so (*how*).

I.



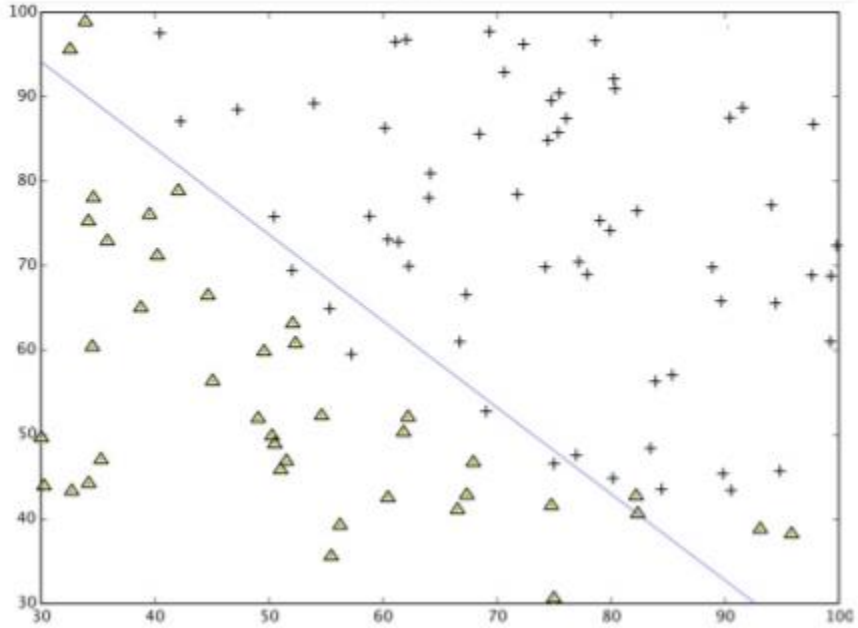
II.



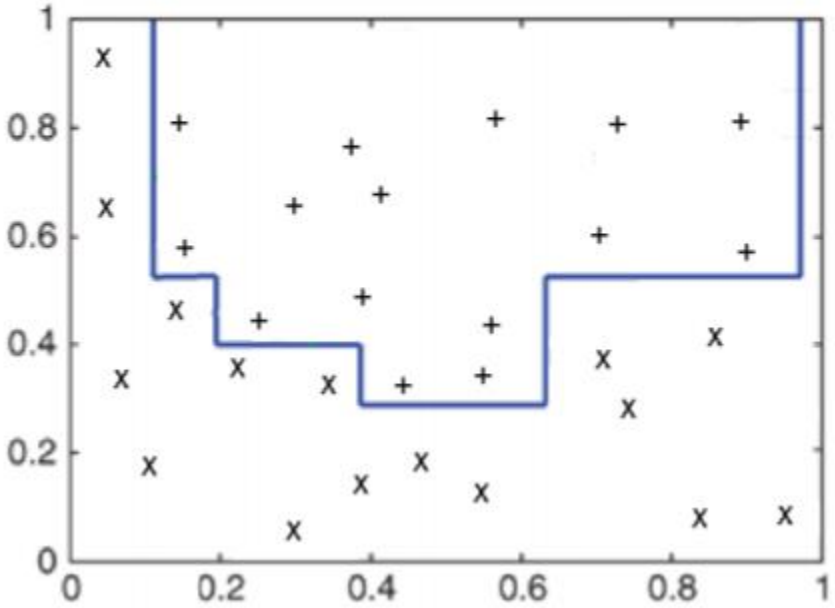
(a) **(3 points)** What predictor(s) produces the decision boundary from figure I and how?

(b) **(3 points)** What predictor(s) produces the decision boundary from figure II and how?

III.



IV.



(c) **(3 points)** What predictor(s) produces the decision boundary from figure III and how?

(d) **(3 points)** What predictor(s) produces the decision boundary from figure IV and how?

Short Answers

What are the problems with fitting a neural network which has multiple layers?

What is one solution that helps to fit these “deep” models?

Short Answers

Why is accuracy an imperfect measure of model quality? Give an example of a model which has high accuracy, but is not desirable.

Short Answers

Other than through bootstrapping, how does a random forest enforce that its composite learners are different?

Short Answers

What is the ROC curve? What are its axes?

Short Answers

What is the difference between pre-pruning and post-pruning in decision trees?