CS 360: Machine Learning

Sara Mathieson, Sorelle Friedler

Spring 2024

HAVERFORD
COLLEGE

Sit somewhere new!

Admin

- Sorelle office hours **TODAY: 4-5pm in H110**
- **Lab 2** due Thursday Feb 8 (week from today)

Outline for Feb 1

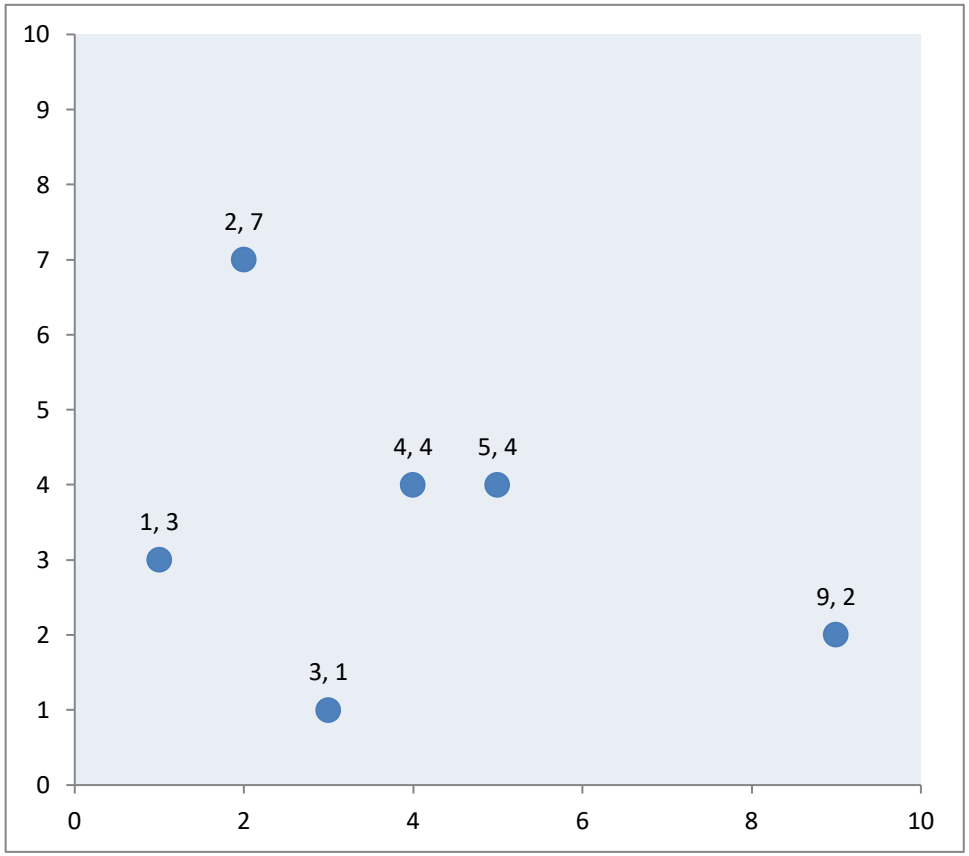
- Finish KD Trees
  - Nearest neighbor algorithm
  - Extending to k > 1
- Evaluation metrics beyond CS260
  - AUC
  - Precision/recall curves

Outline for Feb 1

- Finish KD Trees
  - Nearest neighbor algorithm
  - Extending to k > 1

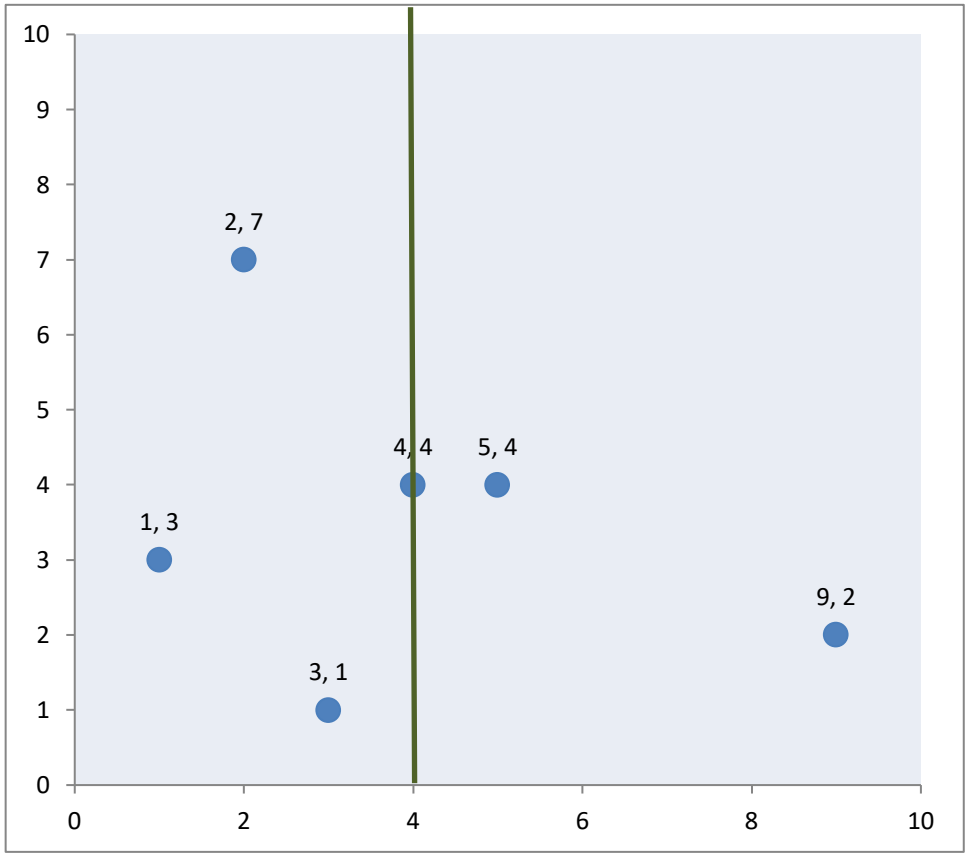
- Evaluation metrics beyond CS260
  - AUC
  - Precision/recall curves

Making a kd-tree



Slide: modified from Sorelle Friedler

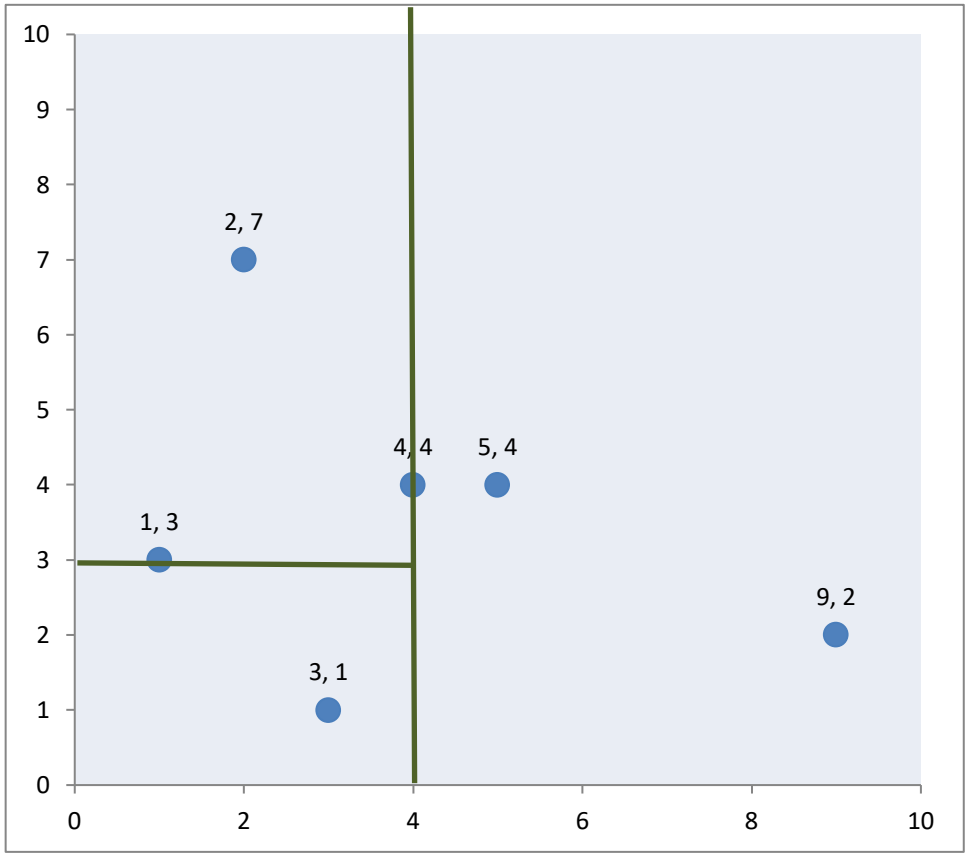
Making a kd-tree

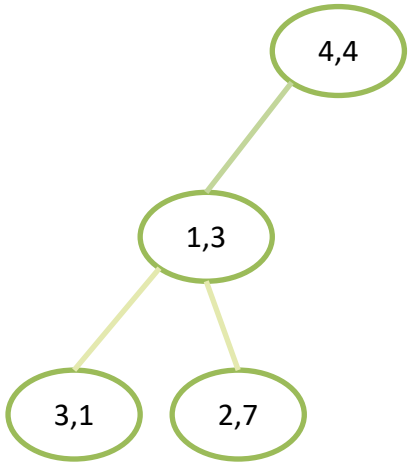


4,4

Slide: modified from Sorelle Friedler

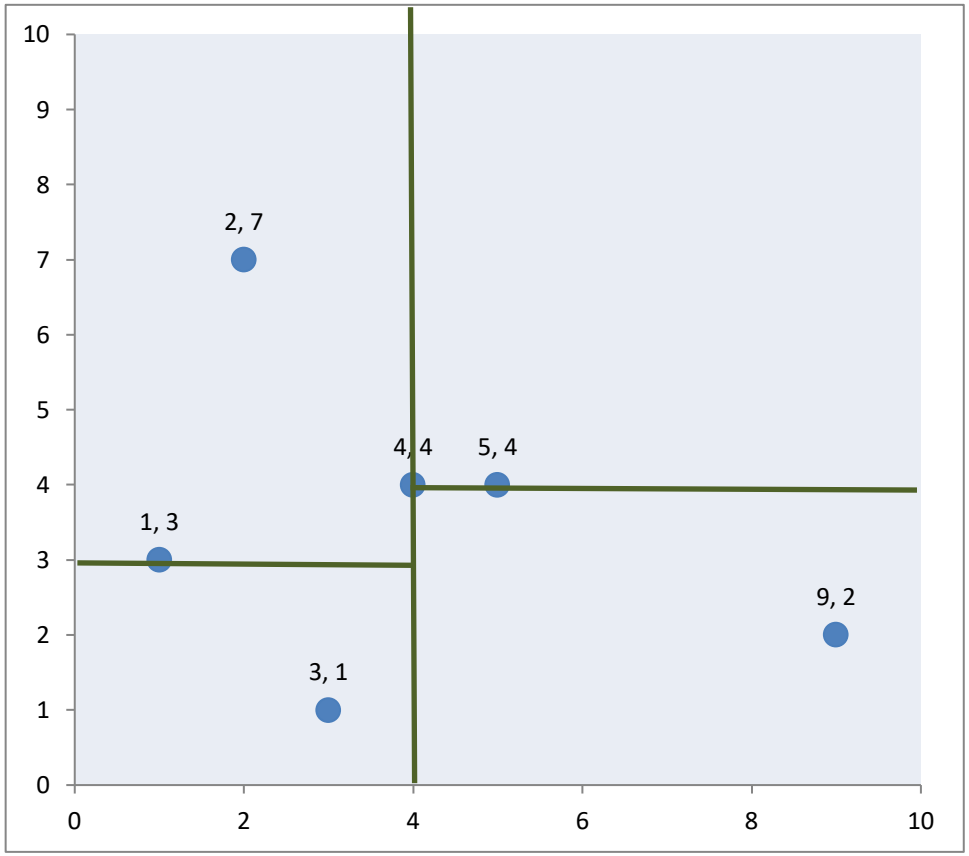
Making a kd-tree

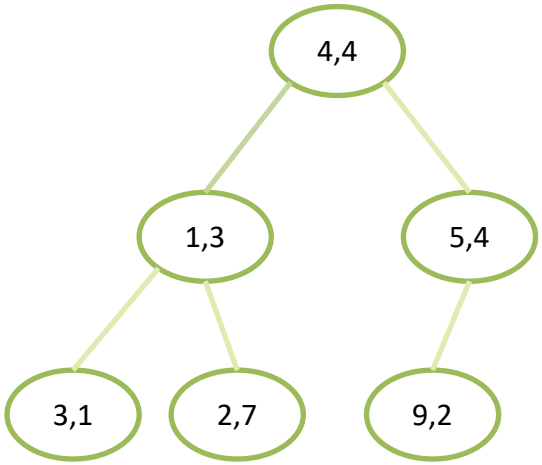




Slide: modified from Sorelle Friedler

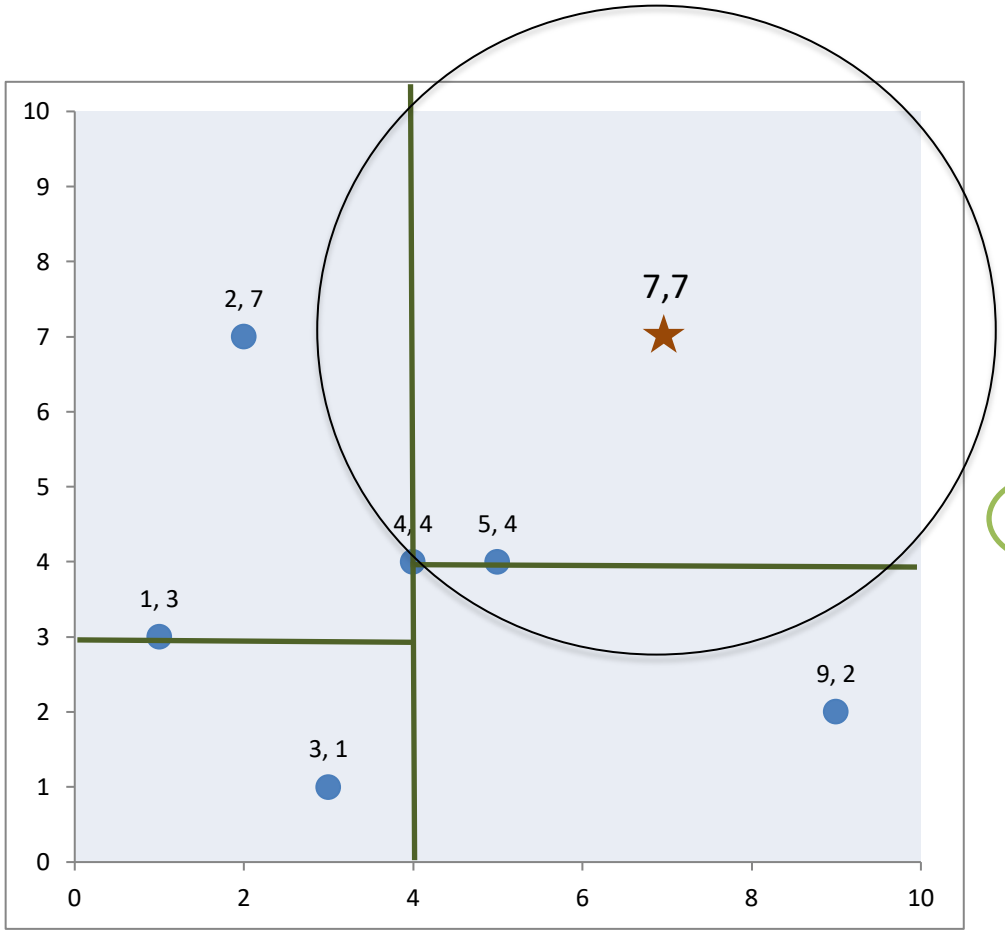
Making a kd-tree

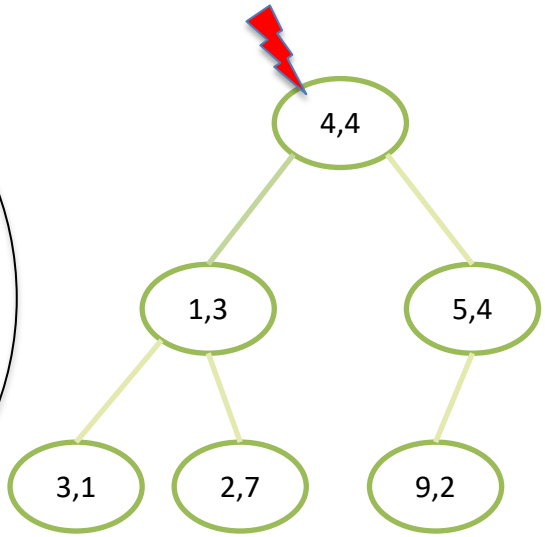




Slide: modified from Sorelle Friedler

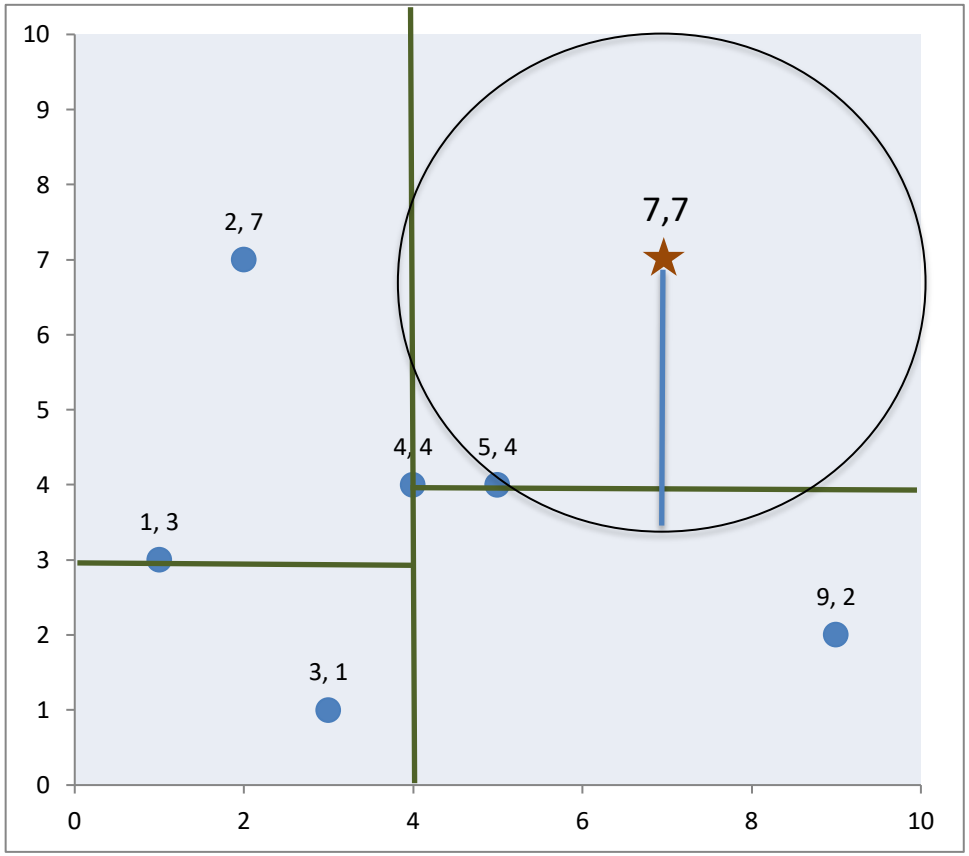
Finding the nearest neighbor

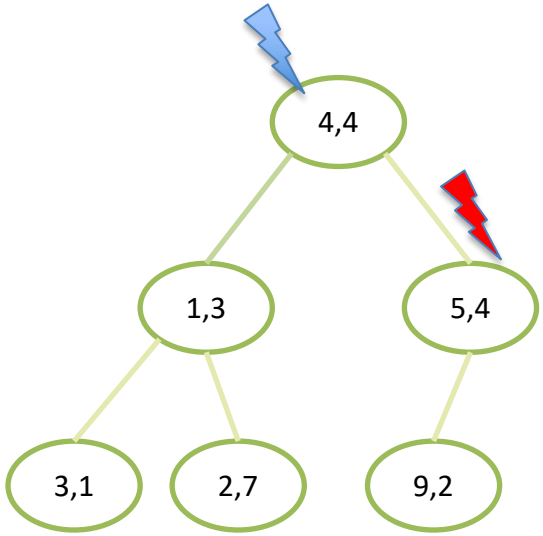




Slide: modified from Sorelle Friedler

Finding the nearest neighbor

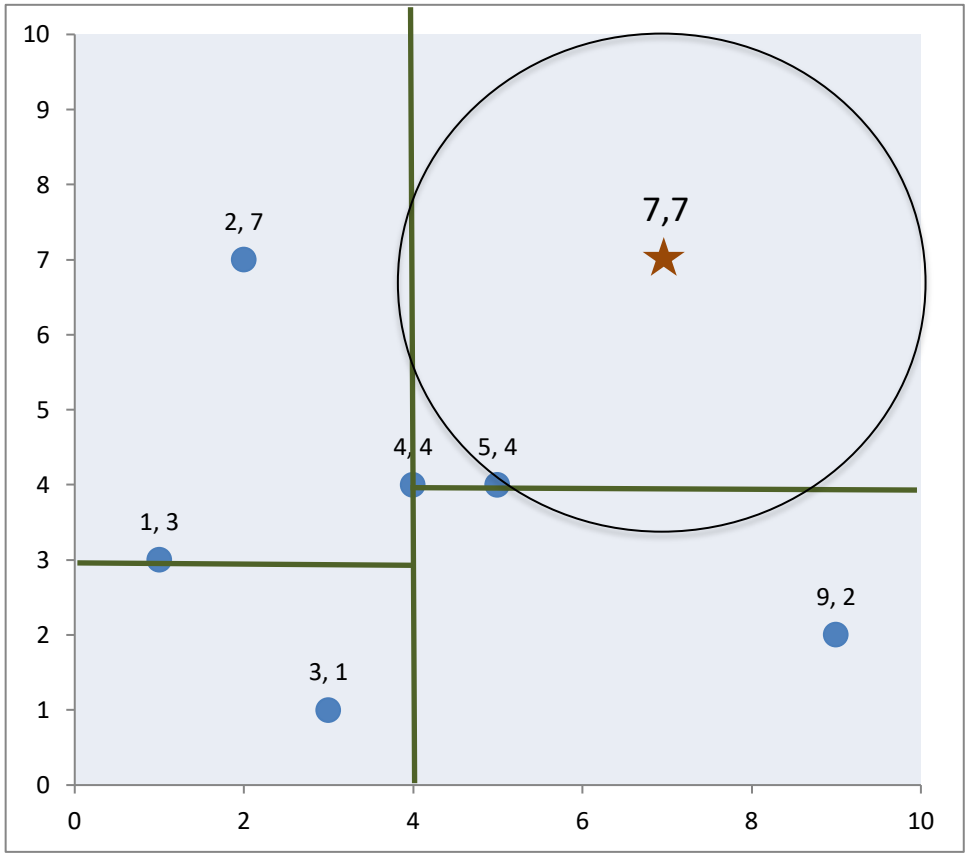


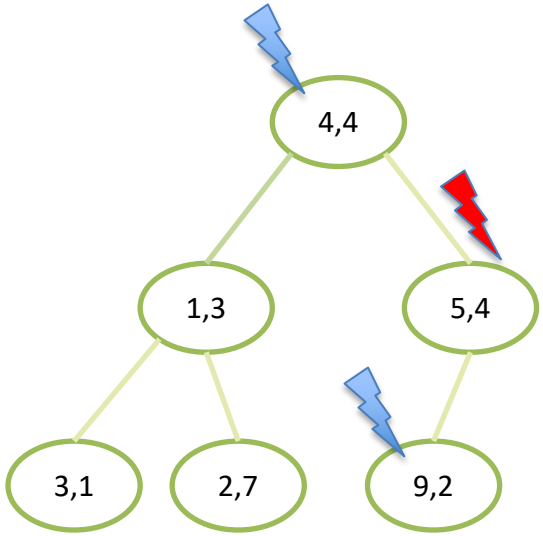


No right child, but need to check left
Due to intersecting hypersphere

Slide: modified from Sorelle Friedler

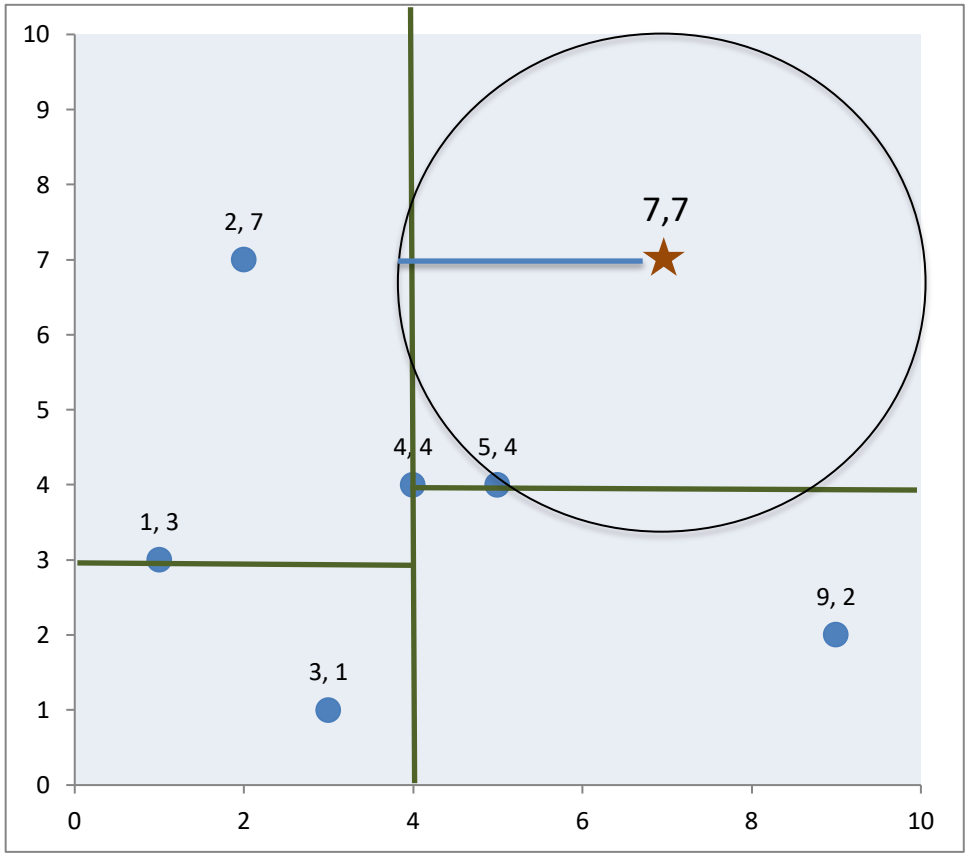
Finding the nearest neighbor

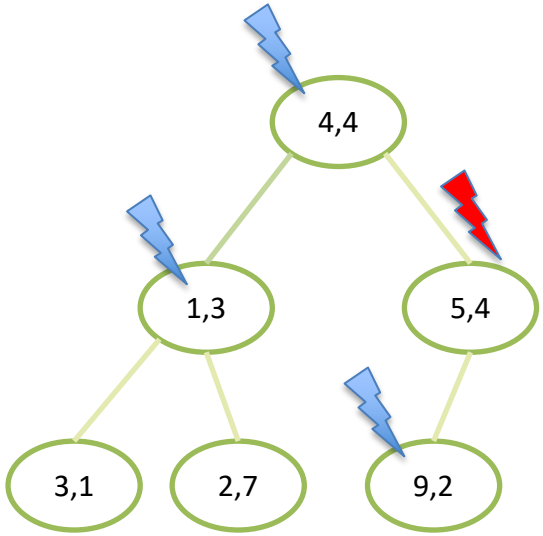




(9,2) is not closer, so keep (5,4) as nn

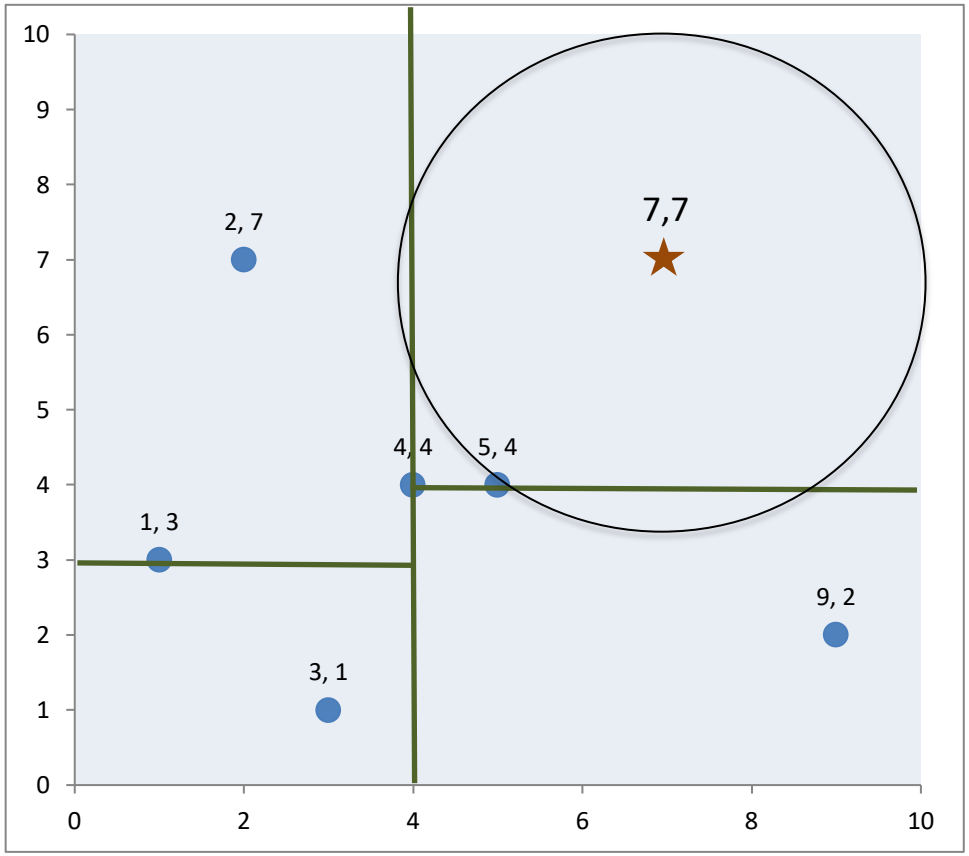
Finding the nearest neighbor

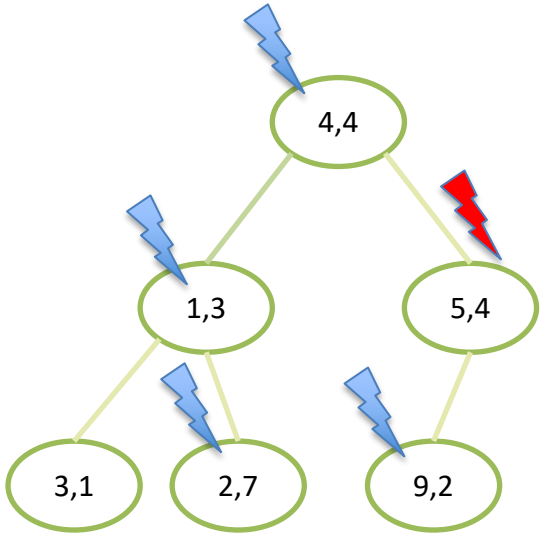




Visit (1,3) due to intersecting hypersphere

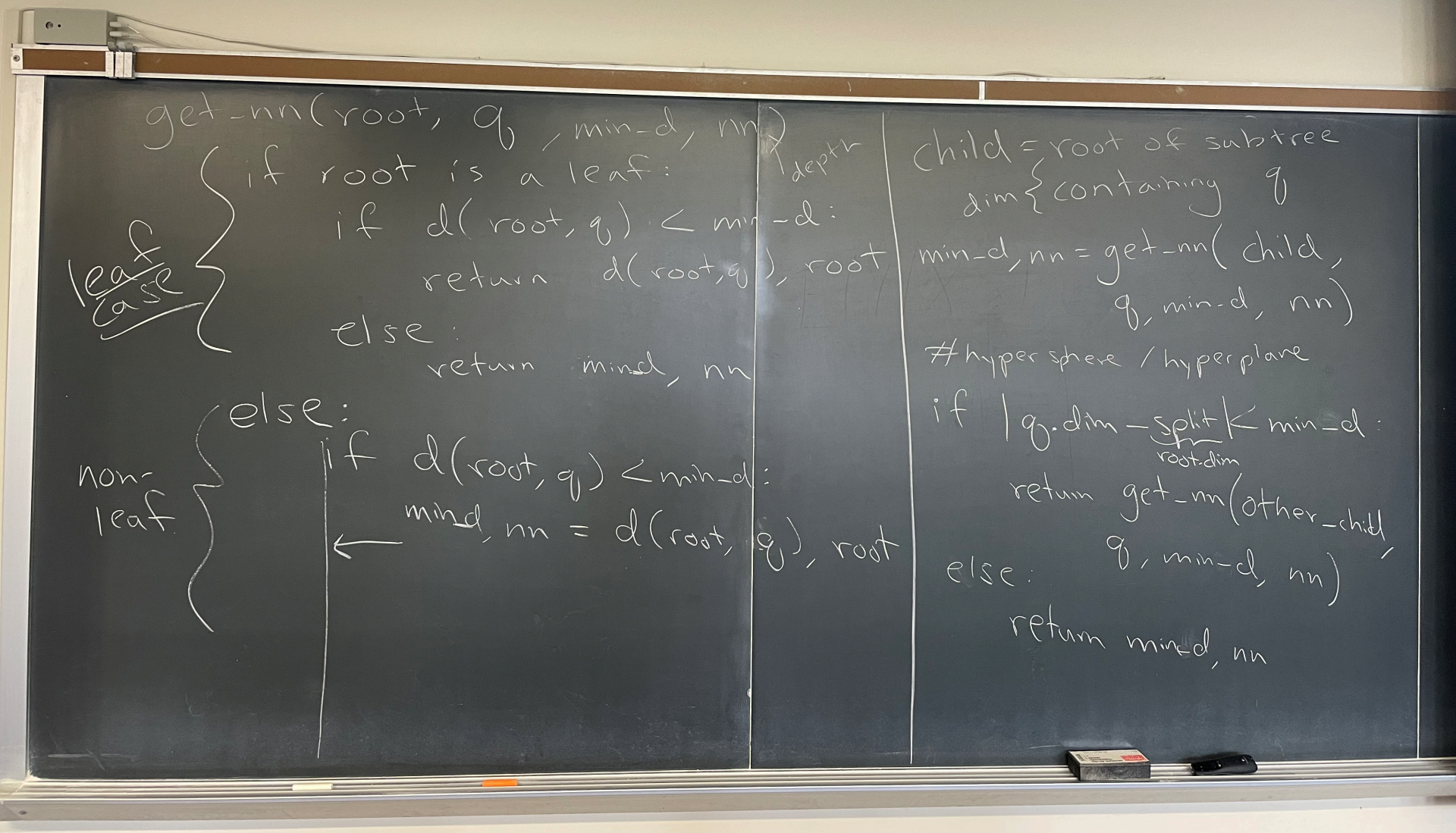
Finding the nearest neighbor





Slide: modified from Sorelle Friedler

Nearest neighbor recursive algorithm

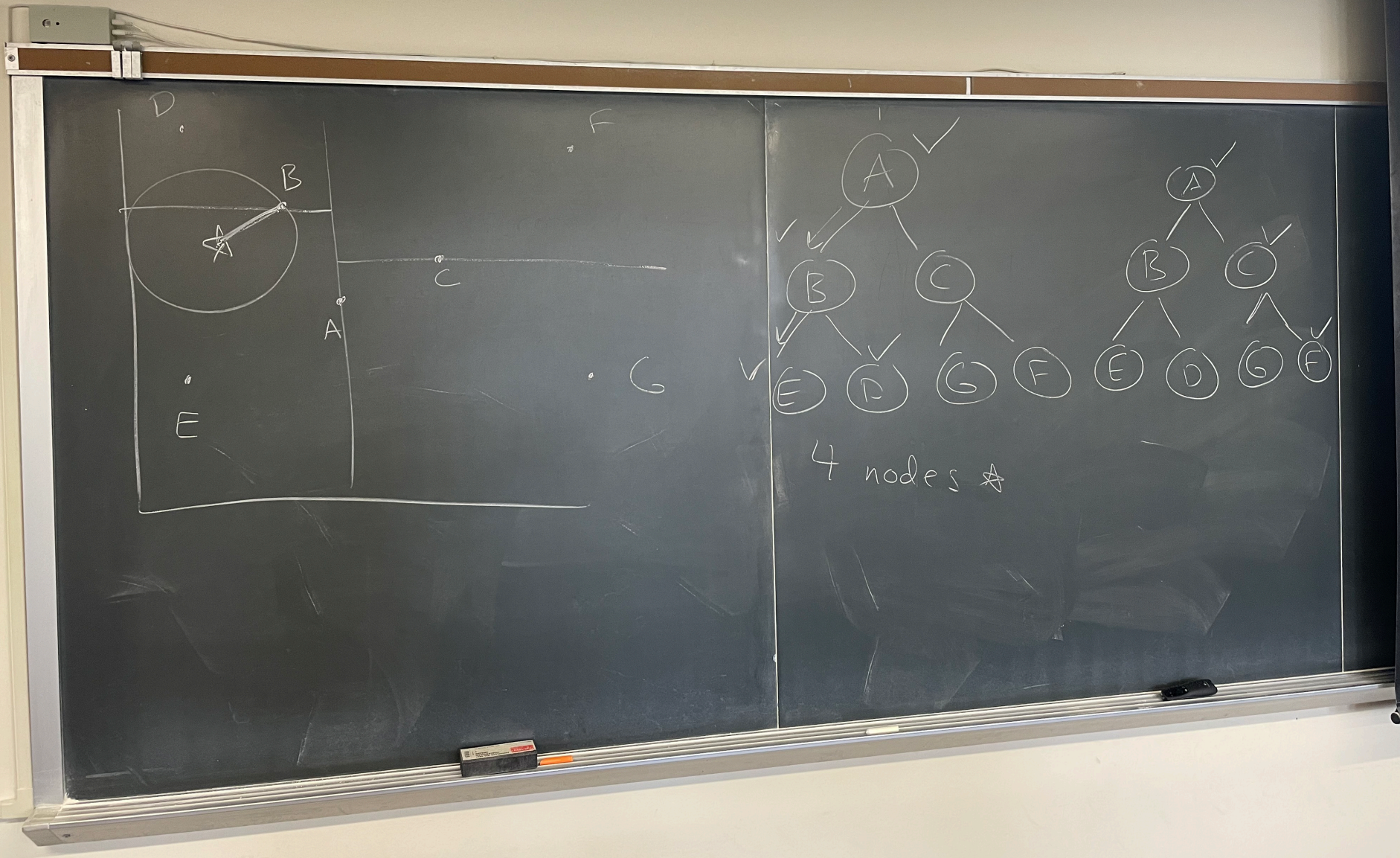


first line

$$\underbrace{\text{dim}}_{\text{cycling}} = \underbrace{\text{depth}}_{\text{increasing}} \ \% \ d$$

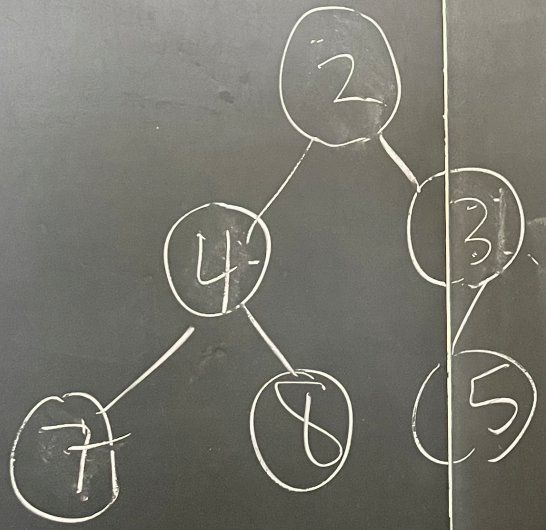
d ↑ total # dimensions

Handout 4



naive ⇒ sort by distance, take top k

distances ⇒ 7, 2, 3, 4, 8, 5, 1

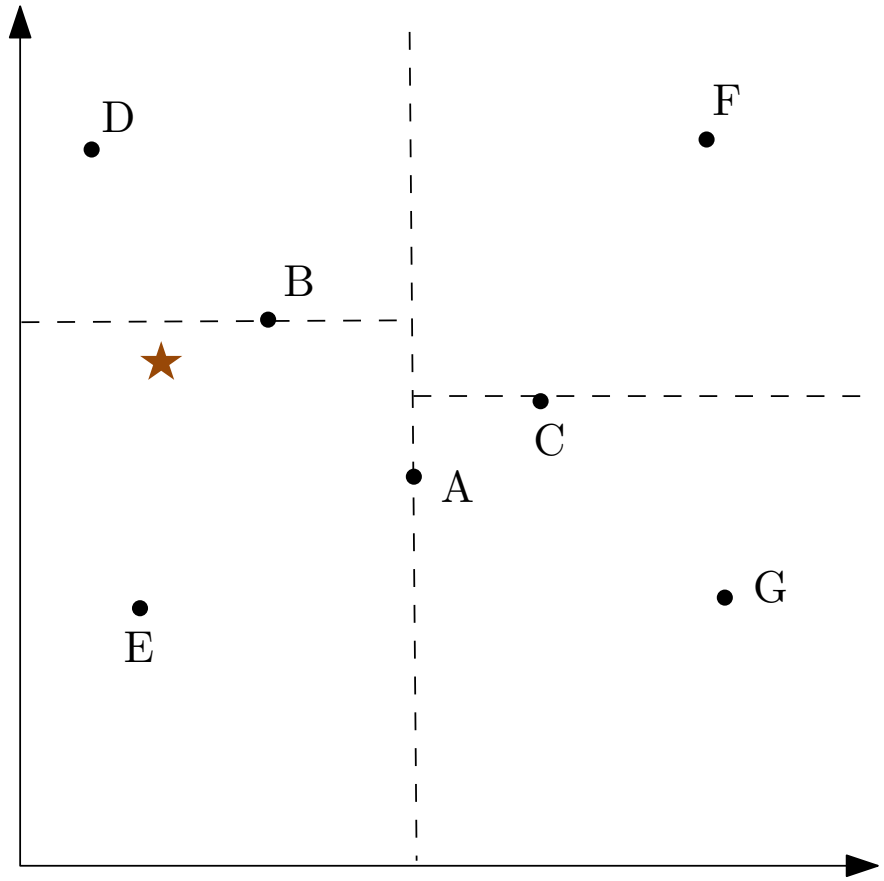


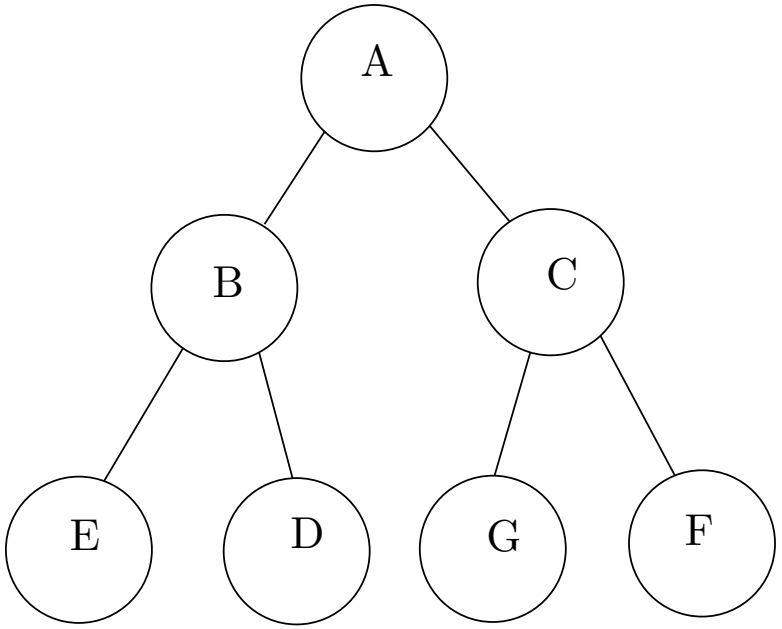
take off 1

take off 2

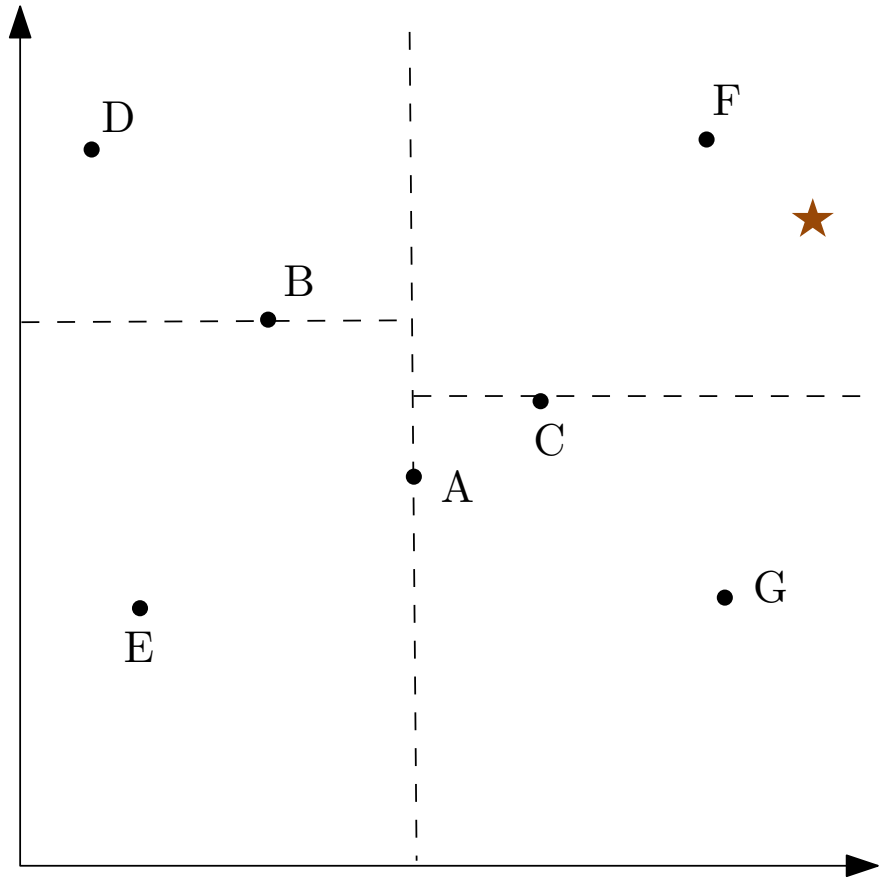
Extending to k > 1

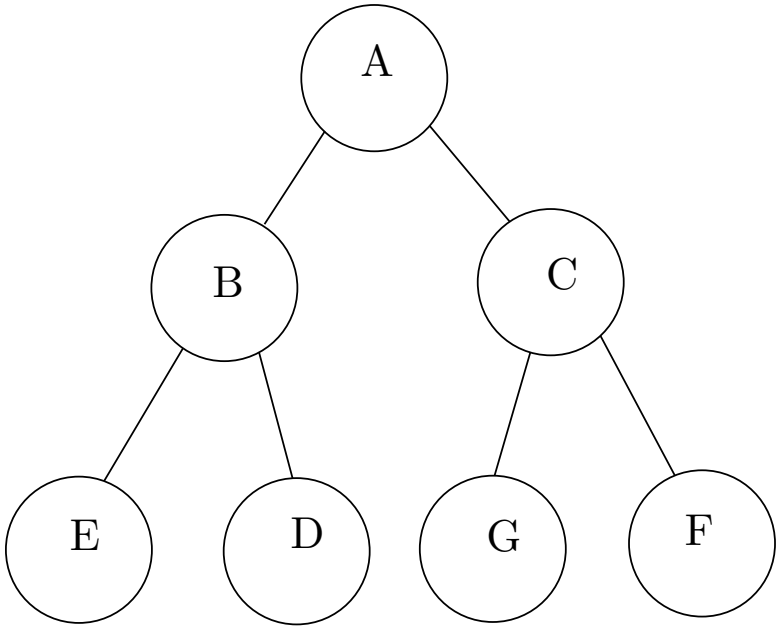
Handout 4, example 1





Handout 4, example 2



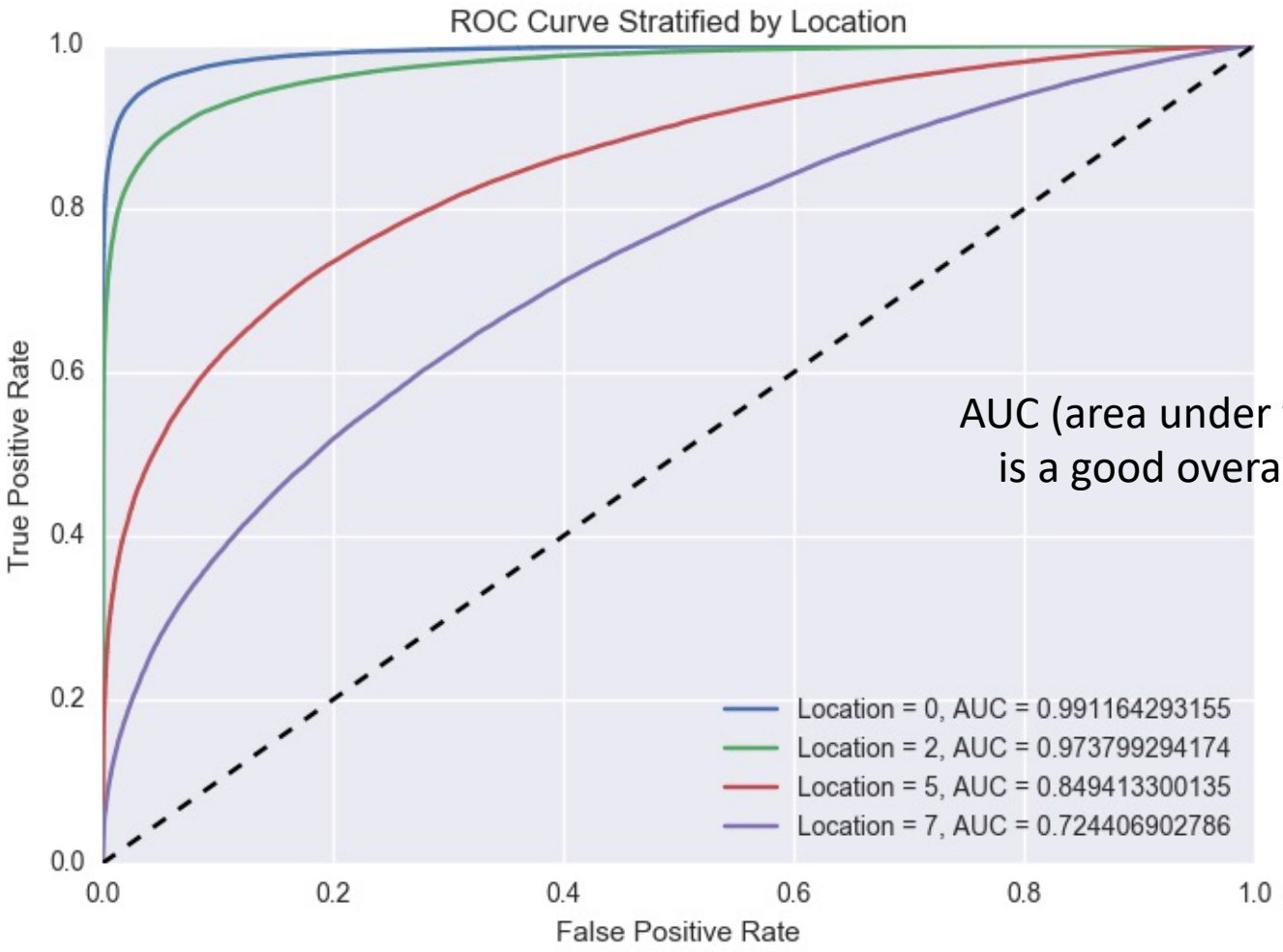


Outline for Feb 1

- Finish KD Trees
  - Nearest neighbor algorithm
  - Extending to k > 1

- Evaluation metrics beyond CS260
  - AUC
  - Precision/recall curves

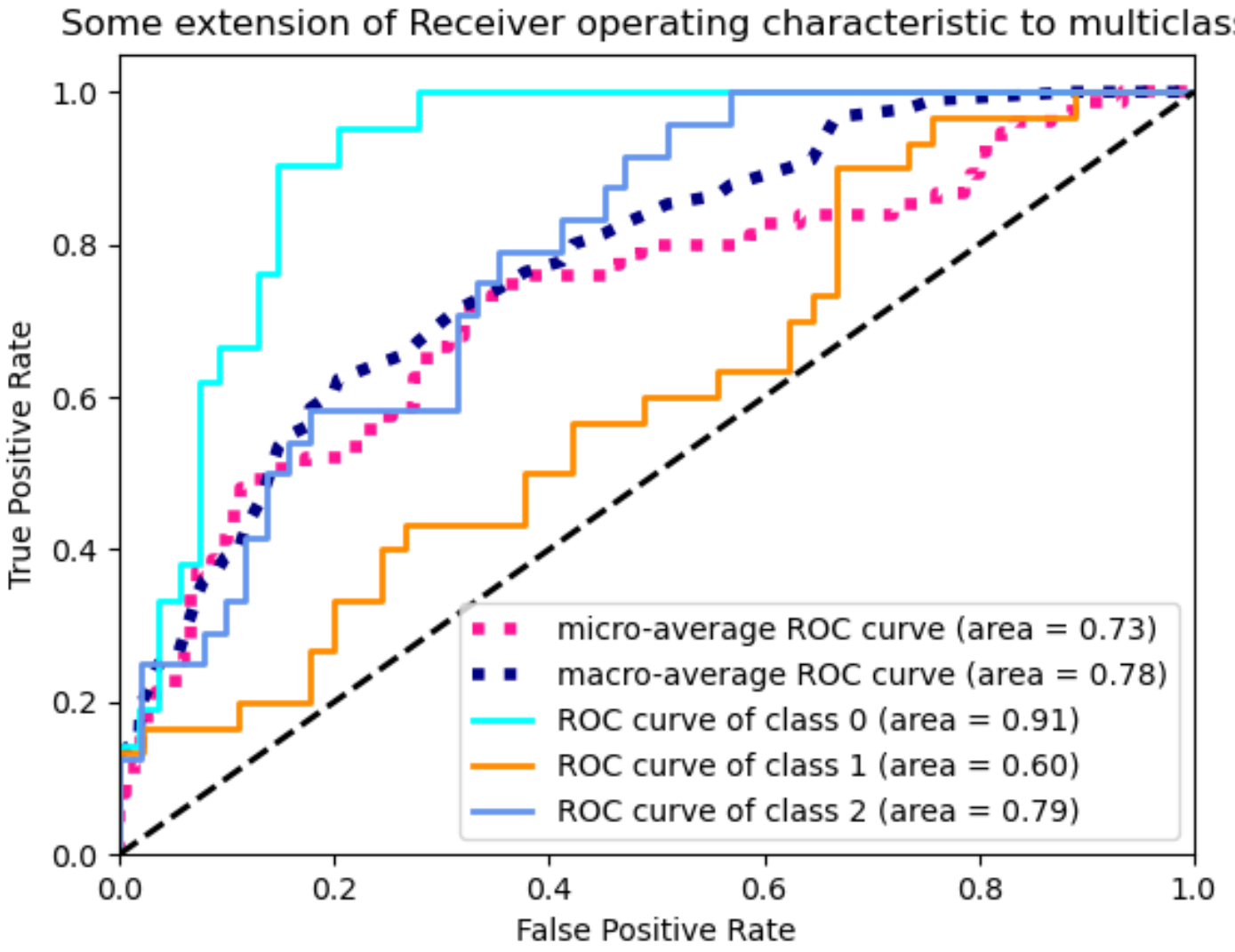
How to compare ROC curves?
AUC (area under the curve)



AUC (area under the curve)
is a good overall metric

Example of a ROC curve from my research
Chan, Perrone, Spence, Jenkins, Mathieson, Song

AUC example



https://scikit-learn.org/1.1/auto_examples/model_selection/plot_roc.html

Precision/Recall curve

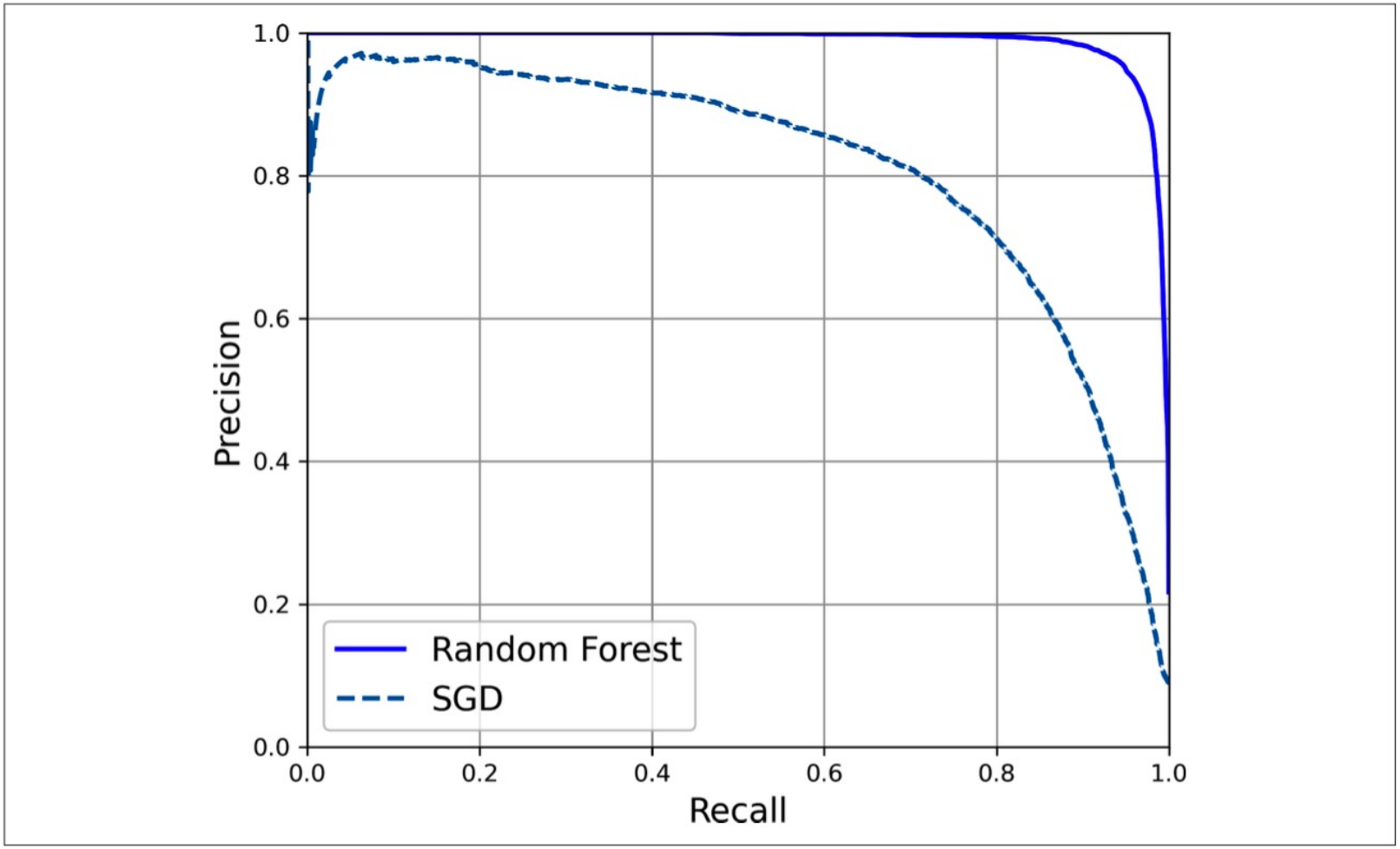


Figure 3-8. Comparing PR curves: the random forest classifier is superior to the SGD classifier because its PR curve is much closer to the top-right corner, and it has a greater AUC

Precision/Recall

example from my research

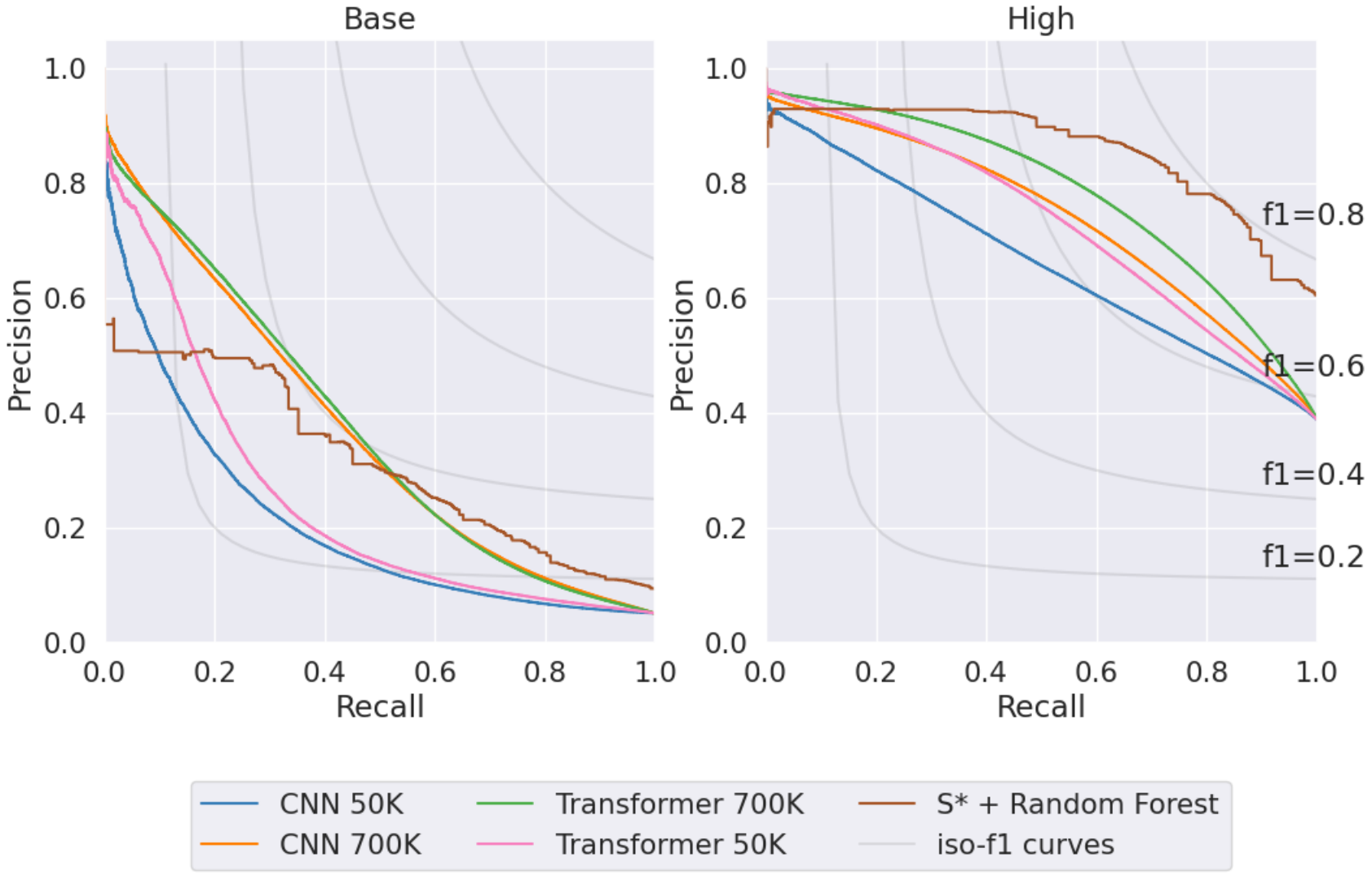
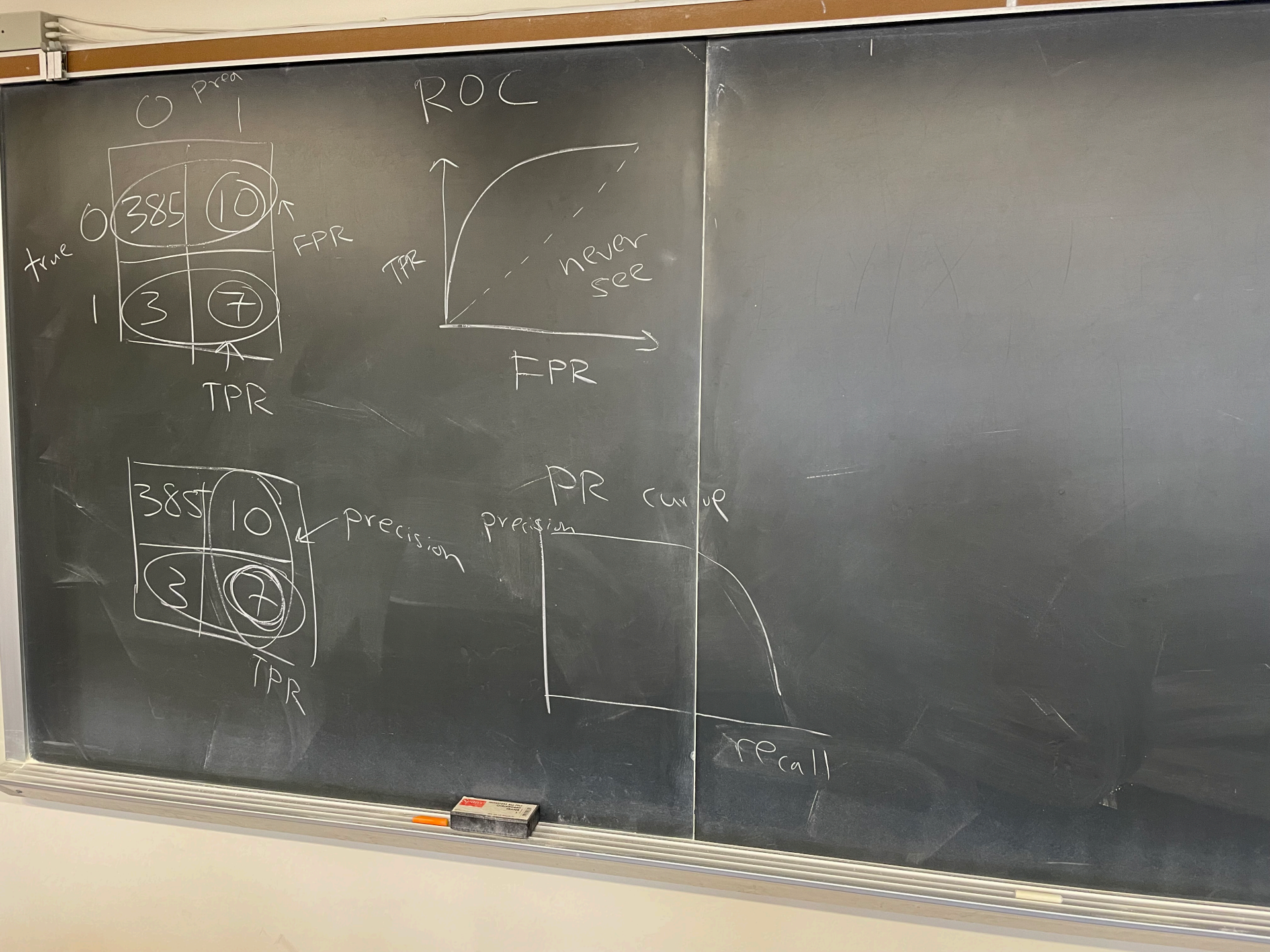


Figure: Jordan Cahoon


0 Pred 1
true 0
385
10
FPR
1
3
7
TPR
ROC
TPR
never see
FPR
385
10
precision
3
7
TPR
PR curve
precision
recall