

6.034 Quiz 3

10 November 2010

Name	
email	

Circle your TA and recitation time **(for 1 point)**, so that we can more easily enter your score in our records and return your quiz to you promptly.

TAs
Martin Couturier
Kenny Donahue
Daryl Jones
Gleb Kuznetsov
Kendra Pugh
Mark Seifter
Yuan Shen

Thu	
Time	Instructor
1-2	Bob Berwick
2-3	Bob Berwick
3-4	Bob Berwick

Fri	
Time	Instructor
1-2	Randall Davis
2-3	Randall Davis
3-4	Randall Davis

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

There are 8 pages in this quiz, including this one, but not including blank pages and tear-off sheets. Tear-off sheets are provided at the end with duplicate drawings and data. As always, open book, open notes, open just about everything, including a calculator, but no computers.

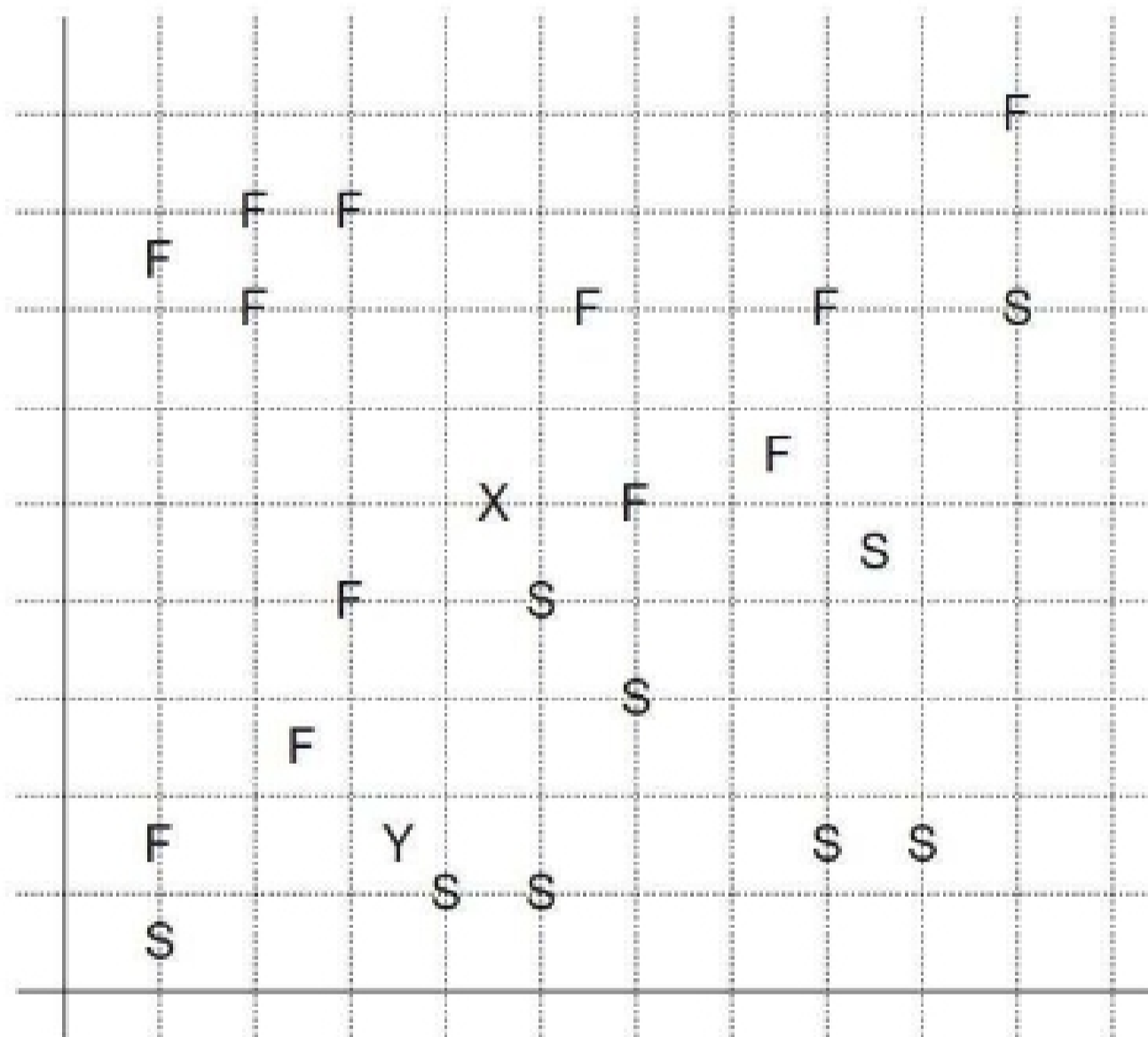
Problem 1: Learning (50 points)

Alice and Bob, a pair of 6.034 students, traveled to DC last weekend for a rally. Over Saturday night, they attended a cocktail party for rally-goers. They knew almost everyone there, but there were a couple of really interesting party crashers.

Alice and Bob decided to use their 6.034 skills to figure out whether the party crashers were at the rally to promote Fear, or restore Sanity.

Part A: Nearest Neighbors (25 points)

During the party, Bob suggests that they look at who the party crashers were spending their time with, given what Bob and Alice know about their friends' reasons to attend the rally (with either Fear/"F" or Sanity/"S").



A1: Draw KNN (10 points)

Alice sketches the above drawing on a napkin, indicating the party crashers, X and Y, and the leanings of their friends, indicated by "F" or "S." She then draws nearest neighbor decision boundaries.

On the above graph, draw the decision boundaries produced by k-nearest-neighbors where $k=1$ and distance measure is Euclidean distance.

A2: More KNN (15 points)

Based on Alice's decision boundaries, what are the classifications for X and Y?

X =

Y =

Alice changes her mind and decides that those boundaries aren't quite right, and tells Bob they should switch to using $k=3$. "Why? That's so hard to draw!" Says Bob. "I think $k=1$ boundaries are too specific," says Alice. What's the name for the problem with $k=1$ decision boundaries?

She decides to classify the party goers using $k=3$. If $k=3$, what are the classifications for X and Y?

X =

Y =

"Okay okay! Based on what you just said, how about $k=21$?" Alice says "I don't think that'd be a good classifier either." What's the problem Alice has with $k=21$?

Part B: ID Trees (25 points)

Alice and Bob give up on classifying the party goers using who they stand near at the party. "Why not look at where they stood during the event?" says Bob. He then pulls up a high-resolution satellite image of the event on his smart phone, zooms and enhances, picks out his and Alice's friends, and sketches all their relative positions on a separate napkin.

Here's the picture he gets. He and Alice argue about the distance their friends were spread out over the event, so he puts in distance from the stage, as well as spread from the center of the mall:

NOTE: lowercase x and y are axes, measuring distance from the stage(y) and the center of the mall(x). There are 16 friends total.