PROFESSOR: Good morning, everyone. Today, we're going to talk about boosting. Boosting is pretty awesome, and it's not as hard as it might seem. It's actually pretty easy, as long as you do it right.

So let's take a look at this boosting problem. Just like with ID trees, you may have noticed there's the ID tree problem where there's a graph, and all of the tests are x and y-axis tests on that graph. And then there's the ID tree problem where there are a lot of crazy different classifiers about characteristics that are discrete. And then all of the sort of ID tree stumps, if you will, are built out of those discrete qualities.

This is an example of a boosting problem of the second type with a bunch of discrete qualities, like evil, emo, transforms, sparkly. And it has an numerical quality number of romantic interests. So it's one of basically the two kinds of boosting problems that you might see.

And now, there is the two-axis Cartesian problem. A good example of that is the hours of sleep versus coffee problem, which I, for one, am planning on doing in tutorial on Monday so that you guys sort of get a sense of both types of problems.

This one looks big. I think it looks-- I mean, it has ten different vampires or non-vampires to classify. It has a whole bunch of possible classifiers. But if you do it right, you can do it fast.

So here's the prompt for the problem. Let's see. After graduating MIT, you get a job working for Van Helsing and Sommers, a famous vampire-hunting consulting agency. Gabriel Van Helsing, one of the two founders, once attended several 6.034 lectures as a guest, and he remembers Professor Winston's vampire identification tree lecture. He assigns you the task of creating a superior classifier for vampires by using boosting on the following data.

So we've got the ID number, which we can use to just write things out in shorthand. We've got the name of several vampires and non-vampires. Then you see whether they're a vampire or not. That's whether their classification is a plus or minus for the quality of vampire.

After that, there's a bunch of possible ways to classify whether they're a vampire or not. There's whether or not they're evil, whether or not they're emo, whether or not they transform, and whether or not they're sparkly, as well as the number of romantic interests that they have.

So for instance, on the one hand, you have Dracula, who's evil, but he's not emo. He can transform into a bat or a cloud of mist. He does not sparkle, and he has five romantic interests-- those three vampire chicks at the beginning, Wilhelmina Murray, and Lucy Westenra.

So on the other hand, you have Squall Leonhart, who's the protagonist of *Final Fantasy VII*, is extremely emo and doesn't have any of the other characteristics. And he's not a vampire. However, he's a nice counterexample for a possible rule that all emo people are vampires because he's very, very emo, and he's not a vampire.

So how will we go about tackling this problem with boosting? Well, there's a whole bunch of different classifiers. And if you think this is all of them-- like evil, emo, transforms, sparkly. romantics interests, and true-- it's actually only half of them. The other half are the opposite versions, but we'll ignore them for now.

So if you look at these, you can probably figure out what they mean-- evil equals yes means vampire is what we're saying here-- except maybe true. You might be saying, why is there one that just says true? The one that just says true says that everybody is a vampire. You might think, oh, that sucks. But it's not that bad since seven of the 10 samples are vampires.

The key, crucial thing about boosting is that for any possible classifier, like classifying on the evil dimension-- which actually sounds like some kind of weird place that you'd go in a comic book. But classifying on the emo dimension or whatever, as long as it's not a 50-50 split of the data, you're guaranteed to be able to use it in some way for boosting.

If there is a 50-50 split, it's like flipping a coin, so it's useless. Because if you had some other thing, like gender equals male or female-- and let's say that was 50-50. It's not. But let's say it was 50-50 between vampire and non-vampire-- it's a useless classifier because it would be just the same as flipping a coin. You'd get no information.

Now, you might say, wait a minute. What about classifiers that get worse than 50-50? What about them? Might not they be even worse than a 50-50 classifier? I claim a classifier that gets less than 50-50 is still better than a classifier that gets exactly a 50-50 split. Is there a

question?

AUDIENCE: Yeah. In the ID tree example, somebody said you used 50-50 classifiers and played around. And after you already produced the elements per set, then you only used 50-50 classifiers [INAUDIBLE] per set.

PROFESSOR: So the question is, in the ID tree example, you might use 50-50 classifiers in later rounds if, for instance, there's a 50-50 classifier except for that most of the things off of that side have been already removed. Let's say there's 20 data points, and there's a classifier that splits it 10 and 10. And it gets half plus half minus on both sides.

But all the pluses from the right side have been removed by some other classifier. You might use it. That's true. But in boosting, you will never use something that's a 50-50 classifier. You never use something that has exactly a 50-50 chance of being correct. Because if it has a 50-50 chance of being correct, it's useless.

And if it has a 50-50 chance of-- no, sorry. Let me specify again. You'll never use something that has a 50-50 chance of giving you the right answer given the weights. That's very, very important. And that may be what your question was getting at. As I'm about to show you and as Patrick told you in the lecture, in later rounds of boosting, you change the weights of each of the 10 data points.

At first, you start with all weights being 1/10. The weights have add up to 1. In this case, you never, ever choose a classifier that gets five of them right and five of them wrong. In the later rounds, you'll never, ever choose a classifier that gets half of the weight wrong-- exactly half of the weight wrong.

But half of the weight may not be half of the data points. So it's possible to choose a classifier that gets half of the data points wrong if it doesn't get half of the weight wrong. And that's similar to an ID tree when you've already gotten things right before. Because you'll see that the weight is going to go to the ones you got wrong.

So I'm not saying that you should throw out right away anything that gets five of the points wrong. Hell, you shouldn't even throw out right away something that gets seven of the points wrong. It's possible-- possible-- that you can get seven of the points wrong, or getting less than half of the weight wrong if those other three points are really, really annoying to get right. And we'll see that later on.

But for insight, at every step along the way, we're willing to choose any classifier that doesn't get 50-50. However, we want to choose the classifier that gets the most of the weight right. By most of the weight, at first, we mean most of the points right. Later, we will mean exactly what I said-- most of the weight.

And if you don't understand that, it's sometimes hard to get it right away when Patrick just lectures through it, introduces a new concept. If you don't understand that, you'll see what I mean when we go through, all right?

So my point I was making before is, what about things that get less than half of the weight right? Well, those are always OK because you can just flip them around, use their inverse, and that gets more than half of the weight right.

It's sort of like-- yeah, it's sort of like my girlfriend always tells me that she is more than 50% likely to choose the wrong direction when you're trying to go between two places, which I'm kind of skeptical of. But I said, if that's really true, then we can just go wherever you didn't say to go, and we'll be more likely to go the right way. So you're actually really good at finding the place that we want to go.

And then she's like, no, that won't work. Because then I'll know that you're going to do that, and I'll double say the wrong way. And then you'll go the wrong way again.

But that not withstanding, you can see you can apply the same concept to boosting. And that's why underneath of this, I have all the opposite versions of all these tests. So what should we be doing to solve this problem more quickly? First, let's figure out which data points are misclassified by each of these classifiers. In other words, if we say all the evil things are vampires and all the non-evil things are not vampires, what do we get wrong?

And if we do that for every classifier, then that'll make it faster later on. Because later on, we're going to go through classifiers, and we're going to have to add up the ones they got wrong. So this chart over here is going to be extremely useful. And on the test that this appeared in, they even made you fill it in to help yourself out.

So let's see. If we said that all the evil equals yes are vampires and all the evil equals no are not vampires, then-- I'll do the first one for you. So we get all of the non-vampires correct because they are all evil equals no. But unfortunately, we get Angel, Edward Cullen, Saya Otonashi, and Lestat de Lioncourt wrong because they are vampires, and they're evil equals

no.

Apparently, Lestat is iffy. But I never read those books, and the Wikipedia article said that in the end, he wasn't that evil. So there we go. Evil equals yes misclassifies 2, 3, 4, and 5.

All right, so let's try emo equals yes. I'll have someone else do it. So let's see if you guys got it. So if we say that all the emo people are vampires and all the non-emo people are not vampires, what do we get wrong?

AUDIENCE: 1, 6, 7, 9.

PROFESSOR: 1, 6, 7, 9-- that's exactly right, and fast. Good. We get 1, 6, 7, and 9 wrong. 1, 6, and 7 are wrong because they are not emo, but they're vampires. 9 is wrong because Squall is emo, and he is not a vampire. Good.

OK, what if we said that exactly the transforming characters are vampires and the ones that do not transform are not vampires? Which ones will we get wrong?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Transforms is the next one over.

AUDIENCE: All of the ones [INAUDIBLE].

PROFESSOR: So which ones would we get wrong if we said that said transforms yes were vampires and transforms no were not vampires?

AUDIENCE: We'd get 8 wrong.

PROFESSOR: We'd definitely 8 wrong because she's not a vampire.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Well, no. It's actually on there.

AUDIENCE: And 3 and 4 we'd also get wrong.

PROFESSOR: Yep.

AUDIENCE: As well as 5.

PROFESSOR: Yes, exactly. OK. Oh, man. You didn't see the chart. You were just like, hmmm. You saw the left. You just said, hm, which one of these are the transforming characters? OK, that's pretty hardcore.

AUDIENCE: [LAUGHTER].

PROFESSOR: But yeah, 3, 4, 5, and 8. No, no, it's definitely given to you. That would be like the worst test ever for international students. Ah, if you don't know these ten characters as vampires, you lose.

All right, so what about that sparkly equals yes is a vampire, and if it's not sparkly, it's not a vampire? This is guaranteed not to go well. What do you think it's going to get wrong?

AUDIENCE: For sparkly?

PROFESSOR: Yeah, sparkly equals yes are the only vampires.

AUDIENCE: [INAUDIBLE] wrong. Angel's going to be wrong. Saya-- so 1, 2, 4, 5, 6, 7, and 8.

PROFESSOR: And 8-- yes, that's right. It gets 1, 2, 4, 5, 6, 7, and 8 wrong. That's pretty awful. But dammit, it gets Edward Cullen right. And he's hard to get correct due to the fact that he's not very much like a vampire. He's more of a superhero who says he's a vampire.

OK, so next, number of romantic interest greater than two. So if they have more than two romantic interest, they're a vampire. And otherwise, they're not a vampire. So which ones would that get wrong? Hm?

AUDIENCE: 3 and 10.

PROFESSOR: Just 3 and 10, that's right. Because Circe had Odysseus. She had Telemachus. Actually, she had that guy she turned into a woodpecker. She had that other guy who was a sea god who caused her to turn Scylla into the nine-headed thing, and probably at least one other person. So Circe it gets wrong.

And it also gets Edward Cullen wrong because he only has one. So 3 and 10. You can tell I thought about this problem when I was writing it up. I wrote this one.

All right, number of romantic interest greater than four. So it's a little bit different this time. Now you have to have at least four romantic interests-- or actually, greater than four, but there are

none that are exactly four-- to be classified as a vampire. Which ones do you think it's going to get wrong?

AUDIENCE: 3, 4, 10.

PROFESSOR: Yup, it is going to get 3, 4, and 10 wrong. Because now, you run into the fact that Saya has that blond [INAUDIBLE] guy, Haji, and Kai. So the last of the positive ones-- because I claim I can derive all the negative ones if you guys give me the positive ones. The last of the positive ones is everybody's a vampire. Who does that get wrong?

AUDIENCE: 8, 9, 10.

PROFESSOR: Yes, OK. Now, I can derive all the negative ones from this without a sweat. Evil equals no-- well, it's 1, 6, 7, 8, 9, 10 without looking at the chart.

Raise your hand if you see why it's 1, 6, 7, 8, 9, 10 without looking at the chart. Raise your hand if you don't. Nobody, OK-- wait. One hand. OK, I saw another one back there too later. They were just more tentative.

OK, it's the complement of A because A is evil equals yes is a vampire. It gets 2, 3, 4, and 5 wrong. So therefore, evil equals no is a vampire is guaranteed to get all the opposite ones.

AUDIENCE: Oh, we could have looked at that too?

PROFESSOR: Yeah, we're looking here, but we're not looking at the big chart there.

AUDIENCE: You can look at any?

PROFESSOR: Oh, yeah. If you can't look at anything, then you're screwed-- unless not only are you as hardcore as this guy, but you've also memorized the numbers. All right, so emo equals no is going to be 2, 3, 4, 5, 8, 10. Transforms equals no is 1, 2, 6, 7, 9, 10. Sparkle equals no is 3, 9, 10.

Romantic interest less than two is everything except 3 and 10-- 1, 2, 4, 5, 6, 7, 8, 9. 1, 2, 5, 6, 7, 8, 9. And then finally, everything but 8, 9, and 10, so 1, 2, 3, 4, 5, 6, 7.

All right, so when we started off, we know what everything gets wrong. I then make a bold claim, because there are n of these, which is 14. I make the claim that there are only six that, in your wildest dreams, you would ever possibly even consider using ever. And the rest, you

would never, ever use. Question?

AUDIENCE: Yeah, I just have a question about the number of romantic interests.

PROFESSOR: Yes.

AUDIENCE: You negated it without an equals on either side.

PROFESSOR: That's true. That works only because there are [INAUDIBLE] 2 or 4. But it should have been negated with a less than or equal. I'm copying off of the quiz. But yes. I noticed that this morning when I was putting myself through my pace. I'm like, there's not a less than or equal to. Wait a minute. And then, oh, wait. It doesn't have any 2's or 4's.

Actually, I don't remember writing them all as 5's and 3's. It's possible that somebody else in the post-editing process changed them all to be about the same number, and then changed the less than or equal tos to be less confusing. It's possible I had Circe at 4, and there was an equal to somewhere. And they were like, forget it. Because I can't think of the fifth romantic interest for her.

So yes, normally, you would have to negate it with an equal to sign, but there happened to not be any things that are equal to 4 or 2 here. So they get away with it this time. But it's good practice.

So I'm claiming that in our wildest dreams, we'd only ever want to use six of these ever. And the other eight, forget it. So let's see. I will call on people at random to-- the first people obviously are getting it really easy-- to tell me which of these you think that you might ever want to use. Give me one you might ever want to use.

AUDIENCE: E.

PROFESSOR: Why, E? Of course. That's the best one. Yes, that's one that you might ever want to use. I'll circle the ones that you might ever want to use. E-- it only gets 3 and 10 wrong. That's amazing. It's like the best class of [INAUDIBLE].

OK, so give me another one that you might ever want to us.

AUDIENCE: F.

PROFESSOR: F. Let's see. F-- F is great. It only gets three wrong. Do people agree that you would ever want

to use F?

AUDIENCE: No.

PROFESSOR: Everyone's saying no. Why not?

AUDIENCE: It's like E, except worse.

PROFESSOR: It's like E, except worse. It's guaranteed at every step, no matter what the weights are, to have a worse accuracy than E. It is definitely good. If E wasn't around, it would be one of our best classifiers of all.

But actually, F is not one of the six. This is why I had them write on the test that there were six, because people might not have found all six. Because people who did figure out not to include F might not have figured out to include some of the ones you want to include. So--

AUDIENCE: I don't understand why you can't use F.

PROFESSOR: Why you can't use F, OK. So we start off with 1/10 weight for all our data points. But let's say during our time of boosting that all ten of the data points have now different weights. So we'll call whatever the weight of 3 is, which you're going to get wrong-- you want to minimize the error, right?

So that weight of 3, which goes into the error of the E, is x. The weight of 10 can be y. So if you're thinking of choosing 3, you know you're-- oh, sorry. If you're thinking of choosing E, your error is x plus y.

AUDIENCE: Sure.

PROFESSOR: If you're thinking of choosing F, your error is x plus y plus z, where z is the error of 4. And since you're never going to have a negative weight, x plus y plus z is always greater than x plus y.

AUDIENCE: You can't choose something without the 3 and the 10 anymore because you already chose E.

PROFESSOR: That's-- yes, you would probably choose something that didn't get the three and the ten wrong. But you would certainly never choose F ever because it's always worse than E. In fact, this is exactly the process that will allow you to find the correct six. And by "will," I mean "can." And by "can," I mean, let's see if you guys get it. Give me another one of the six that you might

keep.

AUDIENCE: K.

PROFESSOR: K is the claim-- sparkly. K, I'm going to say will lose for the same reason as F. It's 3, 9, and 10. It's essentially similar to 3, 4, and 10.

So-- oh, by the way, we should not be only going for the ones with the fewest incorrect. You need to be going for ones that do not have something that is strictly better. In this case, 3 and 10 wrong is strictly better than getting 3, 9, and 10 wrong. Question?

AUDIENCE: Oh, I was going to say transform.

PROFESSOR: You were going to say transform. You are going to be correct. Transforms is one of the ones we need, C. 3, 4, 5 and 8-- there's nothing down here that gets fewer than those wrong. There's nothing that gets us 3, 4, 5 wrong, for instance. Yeah, there's no way to get 3, 4 wrong without getting either 10 wrong, or 5 and 8 wrong.

AUDIENCE: So why is the 8 like that?

PROFESSOR: What?

AUDIENCE: Why not G?

PROFESSOR: Why not G? Why not? Why not G? Let's include G, too.

AUDIENCE: Oh. We don't have to do it all?

PROFESSOR: We need six. No, I just said give me any, and someone gave me the easiest one, E. Question?

AUDIENCE: B

PROFESSOR: Why not B? B looks great. I love B. Let's include B. Does someone else want to give another one that they want to include?

AUDIENCE: A.

PROFESSOR: A. Why not A? Sure. I mean, it's hard to see down here because there might be something better on the bottom. But yeah, there's not. So let's include A. Why not A? I love A. A's great. OK.

So that is now five. There's one more that we need. It's by far the hardest one to find. Find me one more that there's nothing better than it. There's nothing that has a strict subset of the same ones wrong.

AUDIENCE: Question.

PROFESSOR: What?

AUDIENCE: Sorry, can we quickly-- why would you choose A before we chose C?

PROFESSOR: OK, why would you choose A before you've chosen C? Let's say 8 was a real problem for you. And you were just getting-- let's say that 3, 4 and 5, they weren't that bad. They weren't that bad. They weren't that bad. OK, you got them wrong here with transforms. You chose C. But sometime later, 8 was just by far your issue. All right?

3, 4, and 5, and 8-- 3, 4 and 5 are much smaller weights than 8. And then after you've got 3, 4, 5, and 8 wrong, 3, 4, and 5 were still not that bad. And 8 still was a high number.

And then sometime later down the line, you're looking at things. And you're saying, you know what? I really don't want to get 8 wrong again, but I don't mind if I get 3, 4 and 5 wrong. Maybe I'll get it wrong with 2, which I've never gone wrong yet. Actually, none of the ones we've circled here get 2 wrong, so it's probably not that bad to get 2 wrong.

So that's why-- because it doesn't get 8 wrong. If A was 2, 3, 4, 5, 8, you'd never take it. Do you see what I mean? Oh, did someone raise their hand? Did someone find it?

AUDIENCE: I just have a question. You can use the same reasoning for choosing K, right? Because after E, we could have chosen A and said that 9 only a little different--

PROFESSOR: But it's strictly worse.

AUDIENCE: Sorry, sorry, I meant [INAUDIBLE] E to 9 and 10, and then we could have chosen 3, 9, and 10, right? Because--

PROFESSOR: But we chose to use E. Because getting only 3, 10 wrong is better than getting 3, 9, and 10 wrong in any universe. You pick. You see what I mean? It might not be that much worse. It might be only a little bit worse to choose K, but it's always worse. So it-- question?

AUDIENCE: [INAUDIBLE] the right one by reasoning [INAUDIBLE] we need one that doesn't have 3 in it.

Because right now, [INAUDIBLE] get 3 wrong.

PROFESSOR: OK, that's a pretty good insight. What are you thinking about?

AUDIENCE: Well, I'm trying to justify D.

PROFESSOR: You're trying to justify D.

AUDIENCE: [INAUDIBLE].

PROFESSOR: D is huge. It gets more than half of them wrong. But you know what? It gets 3 right. You know what? It gets 10 right. And unlike our things that get 3 and 10 right, which is B, it also gets 9 right. D is the last classifier. You got it.

It's hard to choose one that has seven of them wrong, but D is the last one you might pick. It turns out there's nothing better than this for classifying correctly those annoying data points of Edward Cullen and Squall, and also Circe, who's not that annoying, but she tends to be a problem when romance is concerned.

So these are the six that we might use. We can now ignore the rest of them forever-- or at least until someone reuses this problem or something like that. But we can ignore everything except A, B, C, D, E, G. In fact, why did I even bring that up? All the ones we want are on the front. I'll bring it back down.

Then I'll cross this off with reckless abandon. That even broke off a piece of my chalk. Now, these are the ones that we're actually thinking about using.

There is a chart over here already prepared to do some boosting with these six classifiers, all right? So let's give it a try. Now, remember, in boosting, we always try to choose whichever classifier has the least errors. Is there a question?

AUDIENCE: Sorry, yeah. Before we move on, can you say again a little more slowly what exactly we were looking for when we were choosing our classifiers, like something about the subset?

PROFESSOR: You generally want to take classifiers. So I'll tell you what let's you cross off a classifier. That may be a better way to do it.

You can cross off a classifier as useless if-- and by the way, this is only useful if you can do it faster than just wasting your time looking at all of them. Because if you can't cross off some of

them as useless-- usually on the test, they won't make you. You can just waste your time and have 14 instead of six possibilities every step of the boosting.

But take a look at this-- 1, 2, 3, 4, 5, 6, 7. Then see. Do you have anything here that has a strict subset of these wrong? Oh, look. 2, 3, 4, 5 is a strict subset. This can be crossed off. 1, 2, 5, 6, 7, 8, 9-- anything that's a strict subset? Yes, 1, 6, 7, 9. So it can be crossed off.

1, 2, 4, 5, 6, 7, 8, 9. Let's see. 1, 6, 7, 9 is a strict subset. 3, 9, 10. 3, 10 is a strict subset. 1, 6, 7, 9-- a strict subset. 3, 4, 5, 8 is a strict subset, as is 2, 3, 4, 5. 1, 6, 7, 9 is a strict subset. And up here, 3, 4, 10-- 3, 10 is a strict subset. But the others don't have one, even 1, 2, 4, 5, 6, 7, 8.

In general, you want to keep them. You want to keep every classifier you might use. The only ones you'll never use are ones that there's something else that's just better always by having a strict subset of them wrong.

Hopefully, that was more clear. It's tricky. Very few people realize-- we're brave enough to take sparkly even when it got seven things wrong.

So let's start out some boosting. This wasn't boosting. This was setting yourself up. But it was setting yourself up with the knowledge of how boosting works. Less knowledge, less search. Now we only have to search six things. Ah, I mean more knowledge, less search, not less knowledge, less search.

So we start off with all weights being equal. And since there's ten data points, all ten of the data points are weighted 1/10. OK, we're now weighting all of them equally. Since we're weighting all of them equally, when we want to find the classifier that gets the least error, we just want to find the one that gets the fewest points wrong.

Which one is that? That's our friend E, the first one that people realized was a good one. So we're going to choose classifier E.

What's our error? It's just the sum of the ones we get wrong. So what's our error this time? It's 1/5.

We got points 3 and 10 wrong. They both have a weight of 1/10. 1/10 plus 1/10 is 1/5. So I'll put 1/5, and alpha. Alpha is sort of a vote that will be used at the very end to aggregate our classifier. Alpha is 1/2 natural log of 1 minus the error over the error.

However, I have a little trick for you. It's not that impressive of a trick, but it's a little fun. So since error is 1/2-- sorry, not error-- alpha. Alpha is 1/2 natural log of 1 minus error over error. If error is 1/x, then alpha is 1/2 natural log of x minus 1. That just follows from the math. It's a little shortcut.

If error is in the form of 1/x, then it's just 1/2 natural log of x minus 1, which means since this is in the form of 1/5, everyone, alpha is? 1/2 ln 4. OK, 1/2 ln 4.

So now we come to the part in boosting that many people consider the hardest part, and I'm going to show you how to do it more easily. This is that part where we try to make all the ones we got right, we change their weights to be 1/2. And all of the ones we got wrong, we change their weights to be 1/2.

Here is my automated process. It's called the numerator stays the same method. Here's how it works.

Here's our ten data points. Their current weight is 1/10, all of them. We're about to re-weight them for the next step. So you agree they're all 1/10? They're equal, to start off.

So step one-- erase the denominators. Screw fractions. I don't like them. There's division, multiplication. It's a pain. I just want to add whole numbers. That's what we're going to do.

So which ones do we get wrong? 3 and 10. Circle those. All right. Add the numbers in the circles and multiply by 2. What does that give you? 4. That's the new denominator.

AUDIENCE: Do you always multiply by 2, or just--

PROFESSOR: You always multiply by 2. Add the numbers not in the circles. Multiply by 2. What does that give you?

AUDIENCE: 16.

PROFESSOR: 16. That's the new denominator. The final, crucial step so that we can do this again next round is by far the most mathematically complicated thing here because we have to actually do something with fractions, but it's not too bad-- is we then change everything to be with the same denominator. So the 1/4's become?

AUDIENCE: 4/16.

PROFESSOR: 4/16. All right. I can also uncircle these for next-- ah. I hit the that button. All right. New weights-- 1/16, 1/16, 4/16, 1/16, 1/16, 1/16, 1/16, 1/16, 4/16. Note, the weights add up to 1. The ones you got wrong add up to 1/2. The ones you got right add up to 1/2. You're happy.

So now that you get 4/16 of an error for getting 3 wrong, 4/16 of an error for getting 10 wrong, take a look at these six. I'm not going to call on someone, just whoever's good at math and can add these up more quickly-- just 3 and 10 count as 4. All the others count as 1. Add them up. Tell me which one's the lightest. What did you say?

AUDIENCE: I'd go with B.

PROFESSOR: You'd go with B. It doesn't get 3 wrong. That sounds pretty good to me. Does everyone else like B as well?

I like it. I mean, all of our ones that don't get 3 wrong or 10 wrong, we're only looking at B and D. And D has seven. B has four. So B is the best. B gets 4/16 wrong. Does everyone see that? Because even getting one of 3 or 10 wrong is as bad as all the ones that B gets wrong because of the new weights.

So cool. Let's choose B. That's right. And I sort of gave it away. What's the error that B has? It has four wrong, each of which are worth 1/16. The error is? What? 4/16, or 1/4, whichever is your favorite, which means that the alpha is?

AUDIENCE: 1/2 ln 3.

PROFESSOR: 1/2 ln 3. Bingo. Final round. OK, what did we get wrong? We got 1, 6, 7, and 9 wrong. Oh yeah, we can erase the denominators. All right. What are the numbers in the circles, summed up, multiplied by 2?

AUDIENCE: 8.

PROFESSOR: That's 8-- 1/8. And what about the numbers not in the circle, summed up, multiplied by 2?

AUDIENCE: 24.

PROFESSOR: That's right-- 24, which means I'm going to have to change all the numbers in the circle to 3/24, except I guess I don't because this is the last round. But if I was going to do another round-- let's prepare in case we where, change all of these to 3/24.

Besides, it makes it easier to calculate which one is the best possible classifier because you can just use the numerators and sort of add them up. So while I'm writing that up, you guys figure out which one you like for classifier and call it out to me when I'm done. 3/24. 1/24. 4/24. 1/24. 1/24. 3/24. 3/24. 1/24. 3/24-- wait. I'm off by one here. 3, 1, 4--

AUDIENCE: It's because w1 is not assigned to anything. So w2 is really w1.

PROFESSOR: Aha. You're right. w1 is not assigned to anything, so w2 is really w1 Yeah?

AUDIENCE: There's an extra 1/16 between w1, w2. There's an extra 1/16.

PROFESSOR: Yes, that's true. OK, well--

AUDIENCE: We get it.

PROFESSOR: You get it. H, so what is the best H? You get it because it's right here. See? The process is so foolproof, even a fool like me can get it right while they have the chart wrong. All right, so what's the best classifier?

AUDIENCE: C.

PROFESSOR: You say C. I say that seems pretty reasonable. It only gets 3, 4, 5, and 8 wrong. Does anyone else get a different answer?

AUDIENCE: A.

PROFESSOR: Someone else gets A. I like A. Who said A? A lot of people said A.

Well, let's figure it out. So A gets 1, 5, 6, 7. C gets 4, 5, 6, 7. They're, in fact, equal. Tie-break goes to the lower letter because that's what we said. In fact, I didn't tell you, but that's what we said. Question?

AUDIENCE: So when we were deciding which classifier to use, can we only look at the weights, or do we also have to look at the [INAUDIBLE]--

PROFESSOR: Ignore all previous rounds. The question is, do you only look at the current weights when determining a classifier? Or do you look at the previous rounds as well?

You have to ignore the previous rounds. Trust me. They will be used later in the vote. But it's sort of like tainting the jury a little bit to use the previous rounds when you're doing the current

round. Because you want to start fresh with these new weights, get a new classifier. And then later, everyone will get to make their vote. So you only do it based on the current weights.

AUDIENCE: We don't take any consideration if the last round's 6 was wrong or anything.

PROFESSOR: Nope, although the weights take into consideration is when it's wrong, it's going to increase.

AUDIENCE: OK.

PROFESSOR: Question?

AUDIENCE: Could you theoretically reuse a classifier?

PROFESSOR: The question is, could you theoretically reuse a classifier? Answer-- you absolutely can. When that happens, it essentially gets extra weight because you used it again. But you can never, ever use it twice in a row.

Here's why. Let's say that we want to use-- which was the one we used last over there? B? Let's say we wanted to use B again. What does it give us? 50-50.

If we wanted to use B and then B-- 3, 6, 9, 12 wrong. Always guaranteed to give you 50-50, which is the only way that you can be sure you'll never use it. In fact, that's by design. You could reuse it, but not twice in a row. It could be used later on down the stream.

And it will be used. Because if you do seven rounds, one of them has to be reused. It just gives more weight to whichever one is reused. But yes, A wins against C. C was a perfectly good answer as well. Question?

AUDIENCE: Wait, can you reuse [INAUDIBLE]?

PROFESSOR: What?

AUDIENCE: Instead of A or C.

PROFESSOR: OK. If you could reuse, why doesn't he pick E? E gets eight out of 24 wrong. It's one worse than A and C. That's the only reason. Next step will probably use A-- or sorry. Well, next step, we'll probably use E, frankly-- although maybe not, because we got 3 wrong on A. But pretty soon, we would use E again because E's pretty awesome.

But anyway, here we used A. And we said we got 7/24 wrong. Oh, man, we can't use my little

shortcut. So the answer, it has to be 17/7-- or 1/2 natural log of 17/7. So there we go.

Now, we have to ask, what is the final classifier that we created from all these things? All we do is we sum up all the classifiers we chose. And we multiply them times their weight, alpha. So 1/2 ln 4 times E, whether or not E returns true, plus 1/2 ln 3 times B plus 1/2 ln 17/7 times A, is our final classifier, where E returns a plus 1 if E thinks it's a vampire, and a minus 1 if E think it's not. Same for B and A, all right?

And then we take the sign of this. And I don't mean sine and cosine. I mean just, just is it positive or negative? OK? So the question now on the exam is, how many of the ten data points do you get right if we used this? Let's give it a look see. E is-- so we have romantic interest greater than 2. We have emo yes. And we have evil yes.

So oh my gosh, logarithms, they're sometimes annoying. Do we have to actually add them up? I claim we don't.

Here's a nice special case of having three logarithms on the board. One of two things is true. Either one of those three logarithms is so large that it's bigger than the other two combined, in which case, if that one returns a positive or a negative, it's just positive or negative because that one's big. Or one is not that large, and in which case, any two can dominate the other one, and so is just equivalent to a majority vote.

So I claim we never have to add them when there's only three. You guys see what I mean? Like, let's say one of them was 1/2 log of a billion, and the others were 1/2 log of 3 and 1/2 log of 4. Obviously, whatever the 1/2 log of a billion says, which is multiplied by 1/2 log of a billion, is it's just going to be that, and the others will be ignored.

However, if it's not the case that one of them is larger than the other two combined, then it's a simple vote between the three, because any two can out-vote the other one if they work together. And in this case, let's see, 17/7 is not quite 3. However, log of 4 is certainly not better than log of 3 plus log of 17/7.

It's not even-- log of 4 is equal to log of 2 plus log of 2. And these are both bigger than log of 2. That's rules of logs-- log of 4 equals log of 2 squared, and you can take the 2 out. So these are not big enough that one of them is bigger than the other two combined. So it's just going to be a simple vote.

So let's go through. Dracula-- OK, he's got tons of his little vampyrettes. He's not emo, so E

gets it right. He's not emo. So that gets it wrong. But he is evil. That gets it right. Two out of three vote that he's a vampire-- correct.

Next. Angel-- OK, well, he was in a long running series. He's got plenty of romantic interests. So that gets it right. He's certainly emo. That gets it right. And even though he's not evil, two out of three says he's a vampire, so correct.

Next, Edward Cullen-- well, *Twilight,* here we come. Let's see. He only has one romantic interest, so that gets it wrong. OK. He's emo, so that gets it right. But he's not evil, so two wrong.

So Edward's not a vampire according to our final classifier. But he is. So we got one of the data points wrong. You guys see that? Because two out of three of our classifiers here said that he was not a vampire.

All right, let's see. Saya-- well, she has more than two romantic interests. And she's emo. So even though she's not evil, we get it right. OK?

Let's see. Lestat-- he also has may love interests, is emo, and is not evil. So you get it right. OK, Bianca is evil with many love interests. Even though she's not emo, two out of three, you get it right.

All right, Carmilla-- I'm going to call her Karnstein-- is basically exactly the same as Bianca with the number of romantic interests fixed the way it is. So she will always do the same thing that Bianca does. It's why 6 and 7 always travel together. So we get it right.

Sailor Moon is supposed to be not a vampire. So her number of love interests say that she's not a vampire because she only has one. The fact that she's not evil and not emo says that actually, she's perfectly not a vampire. They all agree. And that's correct.

Squall has only one love interest, Rinoa. And he's not evil, both of which of say he's not a vampire. But he is emo. But two out of three says he's not a vampire. We get it correct.

And Circe, despite her many romantic interests which says she might be a vampire, is neither evil nor emo, and is not a vampire. So we got everything right except Edward Cullen, which perhaps says more about Stephenie Meyers writing than about our boosting.

All right, final question-- Wesley Wyndham, a fellow consultant, has noticed a few correlations

between some of the classifiers you used. He suggests using a new set of weak classifiers consisting of a pair of your classifiers that are logically anded and ored together. For instance, two of the new classifiers would be emo equals yes or evil equals yes, or sparkly equals no and transforms equals yes. So that would cut out Sailor Moon from the transforms cloud.

He believes that you'll be able to classify large vampire data sets-- larger than this one, anyway-- more quickly with fewer rounds of boosting using his system. Do you agree or disagree with Wesley? Explain your arguments.

So this was, you know, the tough concept question. Does anyone have just an instinctual thing other than like, oh, man, it's Wesley. He must be wrong.

AUDIENCE: You'll probably use fewer rounds of boosting because you have more classifiers. But you'll have to search through more classifiers.

PROFESSOR: Aha, that is the rare full-point answer. Very few people realize that Wesley was partially right. They either said something about him being completely wrong, which was wrong, or said that he was completely right.

Yes, it will use fewer rounds of boosting because of the fact that you can-- essentially, one of these boosting already does is sort of gets things to vote together in an [INAUDIBLE] fashion. So it will use approximately half the number of rounds of boosting by being able to combine into two.

But there's a lot of ands and ors. There's in fact n choose 2, where n is the number of vampires. And since using half the number of rounds but taking n choose 2 time for each round is not less time. So that's exactly correct.

Not that many people got full credit on that one because sometimes, they were seduced by Wesley's idea. Or they were just like, it's Wesley. He's wrong-- or just some other funny answer.

Any questions about boosting? Question?

AUDIENCE: How do you know how many rounds of boosting to take?

PROFESSOR: The question is, how do you know how many rounds of boosting to do? The answer is-- so on the quiz, it tells you that you have three. In real life, you might want to just kind of keep it

running until it converges. That's one possibility-- keep it running until it converges to an answer, and it doesn't do anything anymore.

Patrick has a little widget on the 6.034 website, I think, that lets you plunk down some data points and run boosting on them. And you can see that eventually, it converges. The boosting converges to an answer, and it doesn't change.

AUDIENCE: What converges?

PROFESSOR: Basically, the-- not the classifiers you picked, of course, or the weights. But what converges is which ones of your data set you get correct. Because he does his in two-dimensional space rather than like this. And he shows you the lines that boosting is drawing between classification, and colors things in green and red or something like that.

And eventually, it converges where the lines are and which ones it's getting right. It generally converges to getting everything correct. And when that happens, then you can stop.

But that's a good question. And it's not always that easy in the real world. You have to sometimes just say, this is enough for me. I've given it n number of rounds. And that's much more the number of classifiers, so maybe it won't get anything better.