0:0:0.0 --> 0:0:0.130  
Amit Sethi  
Yeah.

0:0:25.520 --> 0:0:26.820  
Amit Sethi  
Ohkay gonna come back?

0:0:28.320 --> 0:0:28.820  
Amit Sethi  
I've come back from.

0:0:31.570 --> 0:0:43.400  
Amit Sethi  
Ohh, how was your session last session interesting about the CNN so they'll be so next Tuesday or they will not give you.

0:0:44.430 --> 0:0:44.930  
Amit Sethi  
Yes, Sir.

0:0:45.70 --> 0:0:46.200  
Amit Sethi  
Talk about and that's.

0:0:47.170 --> 0:0:49.240  
Amit Sethi  
That's almost it, Aaron.

0:0:49.250 --> 0:0:58.160  
Amit Sethi  
Those who are waiting for me to teach advanced ML next semester most likely are not teach it because there's a change in schedule.

0:0:59.170 --> 0:1:5.430  
Amit Sethi  
Now every department is introducing compulsory ML posts for or their underground.

0:1:6.360 --> 0:1:13.850  
Amit Sethi  
Uh, so most likely I'll be teaching that intro not into their mail, but like data science and ML combined for electrical engineering.

0:1:14.200 --> 0:1:19.0  
Amit Sethi  
So again, it will be a basic course which is, uh, like this post.

0:1:19.450 --> 0:1:27.940  
Amit Sethi  
So advance I will teach like one year from with January of 2025 and 17 uh.

0:1:27.960 --> 0:1:37.270  
Amit Sethi  
So point being that this is your chance to pay some attention to other teams and Transformers so that if you want to start doing some projects on that you can.

0:1:37.320 --> 0:1:37.830  
Amit Sethi  
You can do that.

0:1:49.780 --> 0:1:50.680  
Amit Sethi  
The one that question.

0:2:9.920 --> 0:2:10.700  
Amit Sethi  
Then you could find it.

0:3:11.290 --> 0:3:12.910  
Amit Sethi  
But anyone know how to change the background?

0:3:14.340 --> 0:3:16.620  
Amit Sethi  
No, wait.

0:3:16.630 --> 0:3:18.100  
Amit Sethi  
Huh, you know.

0:3:21.850 --> 0:3:23.210  
Amit Sethi  
You have some options inside.

0:3:27.750 --> 0:3:31.300  
Amit Sethi  
So OK, look space.

0:3:36.350 --> 0:3:39.110  
Amit Sethi  
Is this thing ohh?

0:3:47.790 --> 0:3:49.30  
Amit Sethi  
Yeah, this is too smart for me.

0:3:58.80 --> 0:3:58.220  
Amit Sethi  
And.

0:4:2.70 --> 0:4:2.560  
Amit Sethi  
All of it.

0:4:6.310 --> 0:4:7.250  
Amit Sethi  
Ohh thanks.

0:4:12.890 --> 0:4:13.270  
Amit Sethi  
OK.

0:4:13.320 --> 0:4:14.610  
Amit Sethi  
I guess it will have to be this.

0:4:16.800 --> 0:4:25.740  
Amit Sethi  
Ohh, so we were talking about random forest last time and then I wanted to start with this thing, but let's finish the conversation on Monday for us.

0:4:26.590 --> 0:4:28.870  
Amit Sethi  
Umm, So what is that Adam Forest?

0:4:28.880 --> 0:4:32.820  
Amit Sethi  
It's a collection of trees and how do we randomize?

0:4:32.830 --> 0:4:35.10  
Amit Sethi  
We randomize using two things, right?

0:4:35.20 --> 0:4:40.330  
Amit Sethi  
One is a random subsets thing on the random subsets of data.

0:4:47.660 --> 0:4:53.260  
Amit Sethi  
Of data for each you or sample of samples.

0:4:56.430 --> 0:4:57.300  
Amit Sethi  
On each team.

0:5:1.950 --> 0:5:4.80  
Amit Sethi  
OK and 2nd random vibration was.

0:5:6.300 --> 0:5:8.960  
Amit Sethi  
Random subsets of variables.

0:5:13.190 --> 0:5:15.290  
Amit Sethi  
For each, no.

0:5:16.630 --> 0:5:19.20  
Amit Sethi  
OK, for each decision look.

0:5:20.930 --> 0:5:23.440  
Amit Sethi  
For the record, there two randomizations that we did.

0:5:23.750 --> 0:5:25.910  
Amit Sethi  
One was at a tree level, the other was at at node level.

0:5:27.40 --> 0:5:27.350  
Amit Sethi  
Good.

0:5:28.440 --> 0:5:32.650  
Amit Sethi  
So then we can use these to.

0:5:33.340 --> 0:5:38.150  
Amit Sethi  
Ohh basically get some nice things from these random folders.

0:5:38.160 --> 0:5:43.0  
Amit Sethi  
OK, so first thing that we can get is a good idea of generalization.

0:5:44.830 --> 0:5:45.730  
Amit Sethi  
It is generalization.

0:5:53.470 --> 0:5:55.300  
Amit Sethi  
When we say model with.

0:6:1.200 --> 0:6:16.940  
Amit Sethi  
20 bucks 7 and pass it so we're generalization is that there is the gap between training error, training accuracy and test accuracy is not much plus the casing.

0:6:17.900 --> 0:6:20.540  
Amit Sethi  
So we can get a good idea of what the generalization is.

0:6:20.880 --> 0:6:21.190  
Amit Sethi  
Why?

0:6:21.200 --> 0:6:23.20  
Amit Sethi  
Because what we can do is we can.

0:6:23.410 --> 0:6:24.510  
Amit Sethi  
It's ohh.

0:6:24.780 --> 0:6:26.340  
Amit Sethi  
For each three we can find.

0:6:26.350 --> 0:6:26.560  
Amit Sethi  
What?

0:6:26.690 --> 0:6:29.410  
Amit Sethi  
What are the OB samples?

0:6:34.960 --> 0:6:41.20  
Amit Sethi  
If what is OB out of back samples for each week because we are training each tree on a different subset of data.

0:6:44.390 --> 0:6:46.650  
Amit Sethi  
So some of them will not be used for some.

0:6:46.940 --> 0:6:52.410  
Amit Sethi  
So you have 100 peaks and your training on 80% of the data, right?

0:6:52.420 --> 0:7:3.650  
Amit Sethi  
So for every 320% of the data will be OK, though you're so now the same data set, same set of data will be for multiple it would take.

0:7:3.660 --> 0:7:9.130  
Amit Sethi  
Any sample will be out of out of that thing for multiple things.

0:7:11.120 --> 0:7:15.860  
Amit Sethi  
So because it is out of back from multiple trees, it acts as the test data from multiple for those things.

0:7:17.180 --> 0:7:21.50  
Amit Sethi  
So now if you average your decision over those trees, right.

0:7:22.190 --> 0:7:28.300  
Amit Sethi  
For which a given sample is out of that, then you get a good idea of what the generalization error will be, right?

0:7:28.310 --> 0:7:38.470  
Amit Sethi  
So you don't need to set aside validation data, you can just get the set, pick this get an idea of what the validation accuracy will be, while we would be HC.

0:7:39.770 --> 0:7:39.970  
Amit Sethi  
Yeah.

0:7:40.980 --> 0:7:44.260  
Amit Sethi  
So that is one thing that you can get from random for us.

0:7:45.320 --> 0:7:47.790  
Amit Sethi  
2nd is a guarantee against overfitting.

0:7:47.840 --> 0:7:49.890  
Amit Sethi  
So because you are looking at OB.

0:7:56.320 --> 0:8:6.600  
Amit Sethi  
So this avoiding overfitting is only when you look at, not when you look at accuracy of a sample across all trees, only across those trees where it was not teams.

0:8:7.780 --> 0:8:15.450  
Amit Sethi  
So here if you look at the accuracy of a sample then it will give you a good idea because it will give you a good idea of generalization.

0:8:15.460 --> 0:8:21.510  
Amit Sethi  
It's also give you basically tell you whether it is overfitting on that sample or not.

0:8:22.980 --> 0:8:23.570  
Amit Sethi  
OK.

0:8:23.880 --> 0:8:33.310  
Amit Sethi  
Ohh, third thing is ohh something that was proposed here by Freeman and now it has become standard is to get future involvements.

0:8:41.490 --> 0:8:41.630  
Amit Sethi  
Yeah.

0:8:41.970 --> 0:8:43.860  
Amit Sethi  
So how do you get feature importance?

0:8:44.20 --> 0:8:49.680  
Amit Sethi  
Basically, first you train on the trees, OK green the other.

0:8:54.710 --> 0:8:56.70  
Amit Sethi  
Then for Nutan variable.

0:9:0.420 --> 0:9:3.840  
Amit Sethi  
By Mukul feature what does permute mean?

0:9:6.430 --> 0:9:7.230  
Amit Sethi  
How do you convert it?

0:9:9.320 --> 0:9:15.520  
Amit Sethi  
You look at all the samples, right and it has and you look at 1.1 feature.

0:9:16.320 --> 0:9:32.560  
Amit Sethi  
So kasang and randomly permute these values across all the samples and see how much is it decrease in activity but puts the data through this pre training random codes and if there is a large decrease in in accuracy that means the feature was important.

0:9:32.570 --> 0:9:35.280  
Amit Sethi  
We destroyed the feature right, changing its orbit.

0:9:38.170 --> 0:9:39.880  
Amit Sethi  
Uh, so why put muted?

0:9:40.50 --> 0:9:41.540  
Amit Sethi  
Why not give it noisy like things?

0:9:49.810 --> 0:9:52.780  
Amit Sethi  
You can cut the feature by engaging noisy right?

0:9:53.710 --> 0:10:7.160  
Amit Sethi  
So the reason we both permutations because when we are splitting the trees right there, splitting the decision boundaries at each node that is sensitive to the order, it's not sensitive to how the details is.

0:10:8.130 --> 0:10:9.940  
Amit Sethi  
So means subjected to the.

0:10:10.520 --> 0:10:17.950  
Amit Sethi  
Remember we said that at most you can mail d \* 10 -, 1 set of particular note the time minus one comes from the order.

0:10:18.910 --> 0:10:27.630  
Amit Sethi  
It's OK make the values that we have for any particular deliver between those are from, so they're the order.

0:10:28.930 --> 0:10:39.890  
Amit Sethi  
It doesn't matter if the distribution of that behavior is concentrated in this part and it is flat and this part it's the order that that's why he lined up orders entries.

0:10:40.300 --> 0:10:43.110  
Amit Sethi  
You, strictly speaking, you don't need to normalize it.

0:10:43.710 --> 0:10:48.110  
Amit Sethi  
Unlike linear conference or your, it's really awful.

0:10:48.120 --> 0:10:56.680  
Amit Sethi  
Normalized every other unit norm, unit standard deviation and and you you don't have to do that because it only depends on the order of.

0:11:2.190 --> 0:11:7.450  
Amit Sethi  
So when you when you the feature and you look at how much is the decrease in accuracy.

0:11:25.750 --> 0:11:34.800  
Amit Sethi  
So when if there's a large decrease in accuracy, that means the feature was important and it's 2 features are correlated, then it will kind of split the day.

0:11:34.900 --> 0:11:42.620  
Amit Sethi  
That importance between the two features because because we are randomly picking features or variations.

0:11:42.910 --> 0:11:49.0  
Amit Sethi  
So sometimes between two correlated features, sometimes this one is making the decisions, sometimes this one is making the decision.

0:11:49.570 --> 0:11:52.250  
Amit Sethi  
So hopefully the importance will be selected.

0:11:58.710 --> 0:12:4.910  
Amit Sethi  
Ohh, and because it is built on trees, it can handle both discrete and continuously.

0:12:6.780 --> 0:12:6.980  
Amit Sethi  
OK.

0:12:9.840 --> 0:12:10.610  
Amit Sethi  
Any questions?

0:12:10.620 --> 0:12:11.380  
Amit Sethi  
Any doubts?

0:12:11.800 --> 0:12:12.420  
Amit Sethi  
Any comments?

0:12:16.880 --> 0:12:17.20  
Amit Sethi  
Yeah.

0:12:19.870 --> 0:12:20.200  
Amit Sethi  
Or.

0:12:20.510 --> 0:12:20.690  
Amit Sethi  
Yeah.

0:12:21.0 --> 0:12:25.320  
Amit Sethi  
And then and then we get you.

0:12:26.350 --> 0:12:26.970  
Amit Sethi  
Yeah.

0:12:27.10 --> 0:12:30.610  
Amit Sethi  
So he if he should do data processing?

0:12:31.770 --> 0:12:32.980  
Amit Sethi  
Ohh, guess we're ready.

0:12:32.990 --> 0:12:35.320  
Amit Sethi  
Playback on this maybe move on to the community.

0:12:36.20 --> 0:12:36.930  
Amit Sethi  
Are you getting?

0:12:36.940 --> 0:12:38.400  
Amit Sethi  
That is important.

0:12:38.920 --> 0:12:41.950  
Amit Sethi  
It's on the we relevance on you say.

0:12:41.960 --> 0:12:43.510  
Amit Sethi  
That you can distinguish them.

0:12:43.650 --> 0:12:46.550  
Amit Sethi  
You don't want to be down before leaving.

0:12:47.20 --> 0:12:47.470  
Amit Sethi  
That would be.

0:12:48.600 --> 0:12:49.300  
Amit Sethi  
Would you tell?

0:12:49.670 --> 0:12:50.90  
Amit Sethi  
Right.

0:12:50.830 --> 0:12:52.780  
Amit Sethi  
So is there perfectly coordinated?

0:12:52.790 --> 0:13:0.900  
Amit Sethi  
Then it's better to remove one of the very problem goes very supernation is .8 or something like.

0:13:0.910 --> 0:13:3.820  
Amit Sethi  
It's like it's not 1990% of that.

0:13:3.830 --> 0:13:4.350  
Amit Sethi  
It's right.

0:13:4.630 --> 0:13:5.670  
Amit Sethi  
So then what do you do?

0:13:5.710 --> 0:13:9.360  
Amit Sethi  
Like you move one feature you face this shortening the right.

0:13:9.530 --> 0:13:12.790  
Amit Sethi  
So lot of this is just based on like the node items.

0:13:15.300 --> 0:13:20.240  
Amit Sethi  
So in that case, maybe it's better to think both of them if they're permission to.

0:13:21.700 --> 0:13:26.30  
Amit Sethi  
Here both of them and the model and then see both of them have at least moderate.

0:13:39.830 --> 0:13:48.300  
Amit Sethi  
Yeah, but once it does, it says about more interpretive wouldn't just just translated.

0:13:48.910 --> 0:14:2.970  
Amit Sethi  
So if you have a decision date highly interpretable, especially if you control the depth of the drink, if the if you let the depth of the tree become too large, then there's so many if then, else statements, right?

0:14:3.30 --> 0:14:11.390  
Amit Sethi  
Each node is an if then else statements that it's hard, it becomes hard to interpret, but for limited depth it's very interpreted.

0:14:12.770 --> 0:14:16.40  
Amit Sethi  
But then if you have a forest then is it interpretable?

0:14:16.630 --> 0:14:17.200  
Amit Sethi  
I don't.

0:14:17.210 --> 0:14:18.440  
Amit Sethi  
It doesn't seem intolerable, right?

0:14:18.450 --> 0:14:22.780  
Amit Sethi  
Because there are all these trees that are making these internal solutions.

0:14:22.790 --> 0:14:24.230  
Amit Sethi  
But then you are taking the average of your.

0:14:26.310 --> 0:14:29.930  
Amit Sethi  
Pickles no longer interpretable after a certain things.

0:14:44.490 --> 0:14:50.520  
Amit Sethi  
Yeah, I recently used random forest for from the certain work that happened and.

0:14:53.240 --> 0:15:1.780  
Amit Sethi  
It gives like a like if you have tabular data sometimes it gives very good performance compared to like deep neural network signal.

0:15:3.850 --> 0:15:8.980  
Amit Sethi  
And it's much faster to train much faster to implement like in at inference time.

0:15:9.510 --> 0:15:11.570  
Amit Sethi  
So if you have tabular data, start with the.

0:15:12.890 --> 0:15:15.780  
Amit Sethi  
I'll try to start your random forest and see what you get.

0:15:15.790 --> 0:15:17.790  
Amit Sethi  
It says it will give you a good baseline performance.

0:15:21.610 --> 0:15:27.430  
Amit Sethi  
And then I last week I was absent because I was I I was gone to and you really.

0:15:28.450 --> 0:15:30.640  
Amit Sethi  
Uh though global technology conference.

0:15:30.750 --> 0:15:35.130  
Amit Sethi  
So there are now libraries available that will run this.

0:15:35.140 --> 0:15:37.300  
Amit Sethi  
A type of a random forest for texting.

0:15:37.310 --> 0:15:42.80  
Amit Sethi  
Boost OK, which I will briefly cover today.

0:15:42.900 --> 0:15:47.390  
Amit Sethi  
Uh, so they train examples also on on GPU's?

0:15:47.400 --> 0:15:50.770  
Amit Sethi  
Not so if you have very large data very large.

0:15:51.480 --> 0:15:53.120  
Amit Sethi  
Ohh for us to make.

0:15:53.130 --> 0:15:58.900  
Amit Sethi  
Then you can train on on GPUs also with their thousands of kids.

0:16:0.180 --> 0:16:4.730  
Amit Sethi  
So for tabular data you can, you should definitely start with some kind of a random process.

0:16:15.0 --> 0:16:15.500  
Amit Sethi  
OK.

0:16:15.570 --> 0:16:17.230  
Amit Sethi  
Any questions before we go to boosting?

0:16:26.170 --> 0:16:27.190  
Amit Sethi  
So we looked at.

0:16:29.910 --> 0:16:30.490  
Amit Sethi  
Uh.

0:16:31.110 --> 0:16:42.830  
Amit Sethi  
Untangled right in ensembles, what we saw that each model is training separately, but then we are averaging their decision the the models don't look at each other.

0:16:45.320 --> 0:16:48.50  
Amit Sethi  
So that's the 1st way of combining models, right?

0:16:48.740 --> 0:16:49.670  
Amit Sethi  
Uh, unstable.

0:16:52.800 --> 0:16:54.910  
Amit Sethi  
2nd that we looked at was a cascade.

0:16:58.40 --> 0:17:2.980  
Amit Sethi  
So in Cascade you have more than one, right?

0:17:3.50 --> 0:17:4.980  
Amit Sethi  
Let's call it H1 hypothesis one.

0:17:6.70 --> 0:17:9.580  
Amit Sethi  
OK, then it divides the data into two parts.

0:17:9.660 --> 0:17:13.720  
Amit Sethi  
Then you have approaches to proxes 3 etcetera, right?

0:17:14.250 --> 0:17:26.930  
Amit Sethi  
So in gasket this hypothesis 2 will depend on hypothesis one, so it they're not independently, you have to completely train H1 and then we can train Edge 2.

0:17:28.790 --> 0:17:33.220  
Amit Sethi  
Ohh, the third thing that we looked at today is boosting.

0:17:36.950 --> 0:17:45.240  
Amit Sethi  
If we look at two different forms of boosting Adaboost, which is adaptive boosting and gradient gradient boosting in boosting.

0:17:45.250 --> 0:17:47.940  
Amit Sethi  
Also, the models are dependent, so here they are independent.

0:17:55.520 --> 0:17:56.990  
Amit Sethi  
So here also there are dependent.

0:17:57.0 --> 0:18:7.0  
Amit Sethi  
So basically you will have hypothesis H1 and based on that you will training the hypothesis H2 and then based on these two you will train hypothesis Edge 3 and so forth.

0:18:11.660 --> 0:18:15.820  
Amit Sethi  
So here this the arrow itself is that dependency structure.

0:18:16.600 --> 0:18:19.390  
Amit Sethi  
OK, the head of the arrow is what is dependent on the.

0:18:28.580 --> 0:18:31.510  
Amit Sethi  
So your if you have edge four, it will depend on edge one Edge 2 edge.

0:18:37.590 --> 0:18:41.460  
Amit Sethi  
And in Ansamble, H1H2H3 don't depend on each other.

0:18:41.470 --> 0:18:47.210  
Amit Sethi  
So you have H1, you have H2, you have H3 so that independent.

0:18:47.260 --> 0:18:58.580  
Amit Sethi  
Then there's one you can parallelize ensemble, but you cannot penalize cascades and boosting random forests in the combination of cascade and and software.

0:18:59.200 --> 0:19:0.430  
Amit Sethi  
The tree is a cascade.

0:19:0.970 --> 0:19:2.170  
Amit Sethi  
The forest is there and something.

0:19:9.560 --> 0:19:16.340  
Amit Sethi  
So in boosting, your final model will have some weights for each of these, right?

0:19:16.910 --> 0:19:23.240  
Amit Sethi  
Ohh so here what are the weights so the weights here are basically if you have a.

0:19:23.710 --> 0:19:35.490  
Amit Sethi  
If you have three different, if you have three different hypothesis, so basically it will be 1 by K times, H 1 + 1 by K \* H two.

0:19:37.110 --> 0:19:40.110  
Amit Sethi  
Plus one by eight times HQ.

0:19:42.10 --> 0:19:42.350  
Amit Sethi  
Yes.

0:19:42.810 --> 0:19:49.160  
Amit Sethi  
So in boosting, the rate structure will, so all the weights are the same in cache in and summer.

0:19:49.640 --> 0:19:57.400  
Amit Sethi  
In boosting, the weight will not be the same, so weights will be A1 times H1 plus A2 times.

0:20:0.370 --> 0:20:0.730  
Amit Sethi  
Edge 2.

0:20:6.960 --> 0:20:10.490  
Amit Sethi  
Plus, Alpha a times HK.

0:20:12.760 --> 0:20:21.10  
Amit Sethi  
So we have to determine the weights, plus we have to determine how to frame these individual classifiers classifiers.

0:20:21.20 --> 0:20:22.500  
Amit Sethi  
OK, whatever it is.

0:20:29.140 --> 0:20:29.440  
Amit Sethi  
OK.

0:20:31.20 --> 0:20:32.310  
Amit Sethi  
So two things you should know.

0:20:32.320 --> 0:20:40.230  
Amit Sethi  
One is wait, what is the weight assigned to indivision model and 2nd is how is the model dependent on other modes?

0:20:41.740 --> 0:20:45.290  
Amit Sethi  
How is one of the models dependent on the other so?

0:20:48.130 --> 0:21:3.380  
Amit Sethi  
In Cascade, each child model is dependent on the payment model in in boosting, each subsequent model is dependent on all the previous models, but we'll simplify it so that it doesn't have to look at all the models.

0:21:3.390 --> 0:21:11.680  
Amit Sethi  
If we look at like some error composite error of all the models combined, it ohh Yep.

0:21:12.740 --> 0:21:13.810  
Amit Sethi  
So I'll switch slides.

0:21:15.850 --> 0:21:16.520  
Amit Sethi  
Hey.

0:21:20.300 --> 0:21:20.780  
Amit Sethi  
Made some.

0:21:25.850 --> 0:21:26.740  
Amit Sethi  
I'll put this on the.

0:21:30.280 --> 0:21:38.710  
Amit Sethi  
Ohh on on the Ohh on on teams as well, but now we'll talk about two types of boosting.

0:21:38.720 --> 0:21:40.670  
Amit Sethi  
One is and it was.

0:21:54.250 --> 0:21:57.490  
Amit Sethi  
So the first type of boosting is Adam boost adaptive boosting.

0:22:2.310 --> 0:22:4.140  
Amit Sethi  
So the setup is as follows.

0:22:4.250 --> 0:22:8.580  
Amit Sethi  
OK, you have the classifier with n -, 1, right?

0:22:8.710 --> 0:22:11.340  
Amit Sethi  
So you have the classifier with n -, 1 models.

0:22:13.310 --> 0:22:22.460  
Amit Sethi  
The and if you give it the I have sample OK, it will be a weighted combination of all the m, -, 1 models.

0:22:25.300 --> 0:22:26.980  
Amit Sethi  
So can we have weighted combination of?

0:22:28.490 --> 0:22:32.950  
Amit Sethi  
Some all J is equal to 1 to K, K, -, 1.

0:22:39.870 --> 0:22:44.740  
Amit Sethi  
Alpha Gene and HG and geofencing.

0:22:49.800 --> 0:22:51.670  
Amit Sethi  
What is written as K in the slide?

0:22:52.390 --> 0:22:52.870  
Amit Sethi  
OK.

0:22:53.0 --> 0:23:1.830  
Amit Sethi  
And we want to minimize, so this is the decision that is made on the XI sample, OK.

0:23:1.900 --> 0:23:3.330  
Amit Sethi  
And we want to minimize the error.

0:23:5.40 --> 0:23:6.730  
Amit Sethi  
So objective is to minimize the error.

0:23:12.920 --> 0:23:18.140  
Amit Sethi  
Or the next classifier when we add the next model we have m -, 1 models.

0:23:18.150 --> 0:23:21.100  
Amit Sethi  
We add the next model and then we want to minimize the error.

0:23:22.30 --> 0:23:23.480  
Amit Sethi  
So what is your next model?

0:23:23.570 --> 0:23:25.480  
Amit Sethi  
So first, let's look at the next model.

0:23:25.950 --> 0:23:27.300  
Amit Sethi  
So your CNN.

0:23:31.140 --> 0:23:33.410  
Amit Sethi  
Of I don't need the.

0:23:38.920 --> 0:23:46.60  
Amit Sethi  
Uh CMO of XI right is going to be a you just take the previous one.

0:23:47.700 --> 0:23:50.340  
Amit Sethi  
See n -, 1 of XI?

0:23:51.390 --> 0:24:2.190  
Amit Sethi  
Well, I guess you have now alpha N which is the next model that I have added date of the next model that I've added and the next model that I have had it's.

0:24:2.280 --> 0:24:3.300  
Amit Sethi  
So imagine that.

0:24:5.90 --> 0:24:9.720  
Amit Sethi  
Our reports, if it is a binary classifier, not bigger but minus 1 + 2.

0:24:11.380 --> 0:24:16.900  
Amit Sethi  
So then I don't need to normalize the as long as the sign goes this way or that way I get the correct decision.

0:24:18.110 --> 0:24:21.310  
Amit Sethi  
I can keep adding models as long as the sign is correct.

0:24:25.20 --> 0:24:26.730  
Amit Sethi  
OK, so your.

0:24:29.260 --> 0:24:30.350  
Amit Sethi  
Your Pi.

0:24:32.160 --> 0:24:35.110  
Amit Sethi  
Belongs to the set minus 1 + 1.

0:24:42.10 --> 0:24:43.750  
Amit Sethi  
That makes the math easier in this case.

0:24:44.750 --> 0:24:47.500  
Amit Sethi  
So what is it that we want to minimize?

0:24:52.30 --> 0:24:53.940  
Amit Sethi  
You want to minimize exponential error.

0:24:58.190 --> 0:25:0.850  
Amit Sethi  
And that is your sum of.

0:25:3.100 --> 0:25:12.90  
Amit Sethi  
I is equal to 1 to N samples OK and you will have to the power minus TI.

0:25:13.430 --> 0:25:15.880  
Amit Sethi  
That's your, uh, OHS?

0:25:17.560 --> 0:25:22.330  
Amit Sethi  
I don't know label OK and then see M of accepting.

0:25:29.400 --> 0:25:29.670  
Amit Sethi  
Here.

0:25:35.140 --> 0:25:37.380  
Amit Sethi  
So what is this?

0:25:37.390 --> 0:25:38.590  
Amit Sethi  
This is plus 1 -, 1.

0:25:43.650 --> 0:25:44.800  
Amit Sethi  
And what is this?

0:25:44.810 --> 0:25:47.830  
Amit Sethi  
This will give you like a + 1 -, 1 decision.

0:25:47.840 --> 0:26:2.220  
Amit Sethi  
But then when you, uh, I guess it will give you the the continuous value, but you look at the sign, if the signs agree then we will get no exponential error if the sign is disagree, we'll get a high loop.

0:26:2.260 --> 0:26:3.10  
Amit Sethi  
Exponential error.

0:26:4.630 --> 0:26:10.250  
Amit Sethi  
Should we have a minus sign in the exponent and then if the two signs are separate then minus minus will become possible?

0:26:12.130 --> 0:26:14.580  
Amit Sethi  
So then we have high loss and we want to reduce that.

0:26:19.680 --> 0:26:19.870  
Amit Sethi  
Yeah.

0:26:25.10 --> 0:26:26.730  
Amit Sethi  
Alright ohh.

0:26:38.160 --> 0:26:46.240  
Amit Sethi  
OK, so now this part we will break it into the current model being trained and all the previous models.

0:26:47.100 --> 0:26:52.100  
Amit Sethi  
It's so we have some of it was equal to 1 to N.

0:26:54.470 --> 0:26:56.140  
Amit Sethi  
You need to the power.

0:26:57.10 --> 0:26:58.350  
Amit Sethi  
Ohh minus.

0:27:1.630 --> 0:27:3.960  
Amit Sethi  
He to the power minus TI.

0:27:6.260 --> 0:27:8.950  
Amit Sethi  
And then we will have CM minus one excel.

0:27:18.960 --> 0:27:25.160  
Amit Sethi  
And then last alpha M hmm of exit.

0:27:29.210 --> 0:27:29.490  
Amit Sethi  
Great.

0:27:29.500 --> 0:27:33.840  
Amit Sethi  
That's just the expansion of of the other.

0:27:35.550 --> 0:27:41.0  
Amit Sethi  
Of those, Singh, everything so you can see where we are going with this.

0:27:41.190 --> 0:27:42.270  
Amit Sethi  
This part is fixed.

0:27:44.750 --> 0:27:45.900  
Amit Sethi  
Right, this part is fixed.

0:27:45.910 --> 0:27:47.140  
Amit Sethi  
We are not training this part.

0:27:47.460 --> 0:27:48.980  
Amit Sethi  
We have to optimize the second one.

0:27:51.70 --> 0:27:54.550  
Amit Sethi  
Here we have to now find something that optimizes the second part.

0:27:54.730 --> 0:27:56.960  
Amit Sethi  
So basically you will have some over I.

0:28:4.190 --> 0:28:7.200  
Amit Sethi  
OK, the fixed part is east to the power minus TI.

0:28:10.580 --> 0:28:12.840  
Amit Sethi  
See and minus one XL.

0:28:15.720 --> 0:28:18.980  
Amit Sethi  
And then power minus TI.

0:28:21.600 --> 0:28:22.0  
Amit Sethi  
I'm fine.

0:28:22.10 --> 0:28:31.310  
Amit Sethi  
I'm let's make and find outside and for and Pi Edge in Excel.

0:28:34.390 --> 0:28:40.180  
Amit Sethi  
Alright, so think of this part as the weight for the sample.

0:28:42.670 --> 0:28:55.770  
Amit Sethi  
OK, so this is your wait for the next time this is equal to wait for the M3 iteration on of the Ayush sample is a notation WIN.

0:28:58.450 --> 0:29:3.590  
Amit Sethi  
WI because it is not a problems we are putting it in parenthesis.

0:29:8.100 --> 0:29:10.330  
Amit Sethi  
So what does this tell us?

0:29:10.340 --> 0:29:15.390  
Amit Sethi  
That if the laws until m -, 1 step was high.

0:29:16.540 --> 0:29:20.230  
Amit Sethi  
If we were not doing very well on exit sample then increase its weight.

0:29:22.120 --> 0:29:22.370  
Amit Sethi  
OK.

0:29:22.380 --> 0:29:23.110  
Amit Sethi  
In the next iteration.

0:29:25.580 --> 0:29:25.840  
Amit Sethi  
Here.

0:29:28.850 --> 0:29:29.30  
Amit Sethi  
Yeah.

0:29:32.660 --> 0:29:33.440  
Amit Sethi  
Matthew give you.

0:29:35.80 --> 0:29:37.660  
Amit Sethi  
A lot of these losses are based on whether the math is here.

0:29:40.920 --> 0:29:51.280  
Amit Sethi  
Ohh, but also the intuition that if everything is connected loss will be do OK if things are wrong then loss will be the directionally it is correct right?

0:29:51.290 --> 0:29:56.570  
Amit Sethi  
So if your if your TI is minus one and your.

0:29:58.740 --> 0:30:3.570  
Amit Sethi  
The entire model gives you a very large negative number, then the loss will be.

0:30:3.880 --> 0:30:7.970  
Amit Sethi  
It will be very large of very far away from the negative bound on the negative side of the.

0:30:12.170 --> 0:30:15.160  
Amit Sethi  
If it is close to the boundary then it will give you medium loss.

0:30:15.210 --> 0:30:18.890  
Amit Sethi  
If it is misclassified and it's far away and the other side, they will give you high boots.

0:30:25.950 --> 0:30:26.620  
Amit Sethi  
Ohh.

0:30:29.800 --> 0:30:35.780  
Amit Sethi  
Alright, so this first part is the weight that you will give to the sample.

0:30:36.890 --> 0:30:46.570  
Amit Sethi  
So the weight of the sample is very individualized to the sample it is based on how much error you are making till the m -, 1 except.

0:30:48.510 --> 0:30:52.830  
Amit Sethi  
OK, if you are making more error on this sample, increase its weight.

0:30:53.270 --> 0:30:56.180  
Amit Sethi  
If you're making less error on this same sample, reduce it.

0:30:59.180 --> 0:30:59.480  
Amit Sethi  
Why?

0:30:59.490 --> 0:31:6.940  
Amit Sethi  
Because this part, if you look at what is in the power, that is exactly the error that you're making till m -, 1 step.

0:31:14.820 --> 0:31:15.450  
Amit Sethi  
You there?

0:31:17.960 --> 0:31:18.250  
Amit Sethi  
Yeah.

0:31:18.260 --> 0:31:19.550  
Amit Sethi  
So there are two ways.

0:31:19.720 --> 0:31:22.630  
Amit Sethi  
One is the sample weight, the other there's some model.

0:31:23.380 --> 0:31:25.630  
Amit Sethi  
Alpha is a model grade, W is the sample.

0:31:28.730 --> 0:31:32.810  
Amit Sethi  
Just then spoke of that they can translate down in this coming to the previous one, yes.

0:31:33.920 --> 0:31:38.170  
Amit Sethi  
So AA N is modeled.

0:31:43.380 --> 0:31:43.690  
Amit Sethi  
OK.

0:31:43.720 --> 0:31:48.520  
Amit Sethi  
And Wim is sampled.

0:31:50.220 --> 0:31:56.940  
Amit Sethi  
So that's why it has an ID and, but it also has A and because it's individual to each stage.

0:31:58.940 --> 0:32:0.640  
Amit Sethi  
Each additional model that you're adding.

0:32:8.600 --> 0:32:8.890  
Amit Sethi  
So.

0:32:10.750 --> 0:32:16.320  
Amit Sethi  
At the emerge step, we can find the weight, but at the previous step, what is the weight at the first step?

0:32:16.330 --> 0:32:17.40  
Amit Sethi  
What is the weight?

0:32:17.520 --> 0:32:23.610  
Amit Sethi  
So by default WI of one is all one by it's constant.

0:32:24.520 --> 0:32:26.840  
Amit Sethi  
All samples are equally weighted when you're first starting.

0:32:41.780 --> 0:32:41.960  
Amit Sethi  
Here.

0:32:47.460 --> 0:32:50.180  
Amit Sethi  
Alright, so what's the optimal alpha M?

0:32:51.850 --> 0:33:1.890  
Amit Sethi  
Optimal sample rate we can find that by differentiating this expression and setting with respect to alpha and setting it to zero. It's.

0:33:7.670 --> 0:33:10.480  
Amit Sethi  
OK, so that's what we'll do.

0:33:10.540 --> 0:33:16.130  
Amit Sethi  
Let's derivation is slightly tricky, but it's not that difficult.

0:33:17.260 --> 0:33:21.100  
Amit Sethi  
We have to separate the terms that depend on. Ohh.

0:33:23.720 --> 0:33:25.440  
Amit Sethi  
That depend on Alfheim versus not.

0:33:29.960 --> 0:33:30.170  
Amit Sethi  
Good.

0:33:32.450 --> 0:33:32.830  
Amit Sethi  
OK.

0:33:34.990 --> 0:33:36.380  
Amit Sethi  
Can I and it is the top?

0:33:47.160 --> 0:33:54.220  
Amit Sethi  
So we will write this as your error is so ohh.

0:33:56.520 --> 0:33:58.200  
Amit Sethi  
Is some more work file.

0:34:0.290 --> 0:34:0.820  
Amit Sethi  
And.

0:34:2.940 --> 0:34:3.930  
Amit Sethi  
Wim.

0:34:5.710 --> 0:34:5.960  
Amit Sethi  
Here.

0:34:6.630 --> 0:34:10.430  
Amit Sethi  
Uh, and then into the power minus.

0:34:13.140 --> 0:34:13.690  
Amit Sethi  
Home.

0:34:18.810 --> 0:34:20.570  
Amit Sethi  
Alpha M, right.

0:34:20.750 --> 0:34:26.320  
Amit Sethi  
And then you have TI and then you have CNN.

0:34:27.50 --> 0:34:28.280  
Amit Sethi  
Ohh not seeing anything.

0:34:31.630 --> 0:34:32.680  
Amit Sethi  
Hmm XN?

0:34:33.210 --> 0:34:40.780  
Amit Sethi  
If so, this summation over all I right we will split it into two parts.

0:34:42.370 --> 0:34:42.530  
Amit Sethi  
OK.

0:34:43.550 --> 0:34:46.210  
Amit Sethi  
Those better we have correct classification.

0:34:46.220 --> 0:34:47.790  
Amit Sethi  
Those where we have missed classification.

0:34:50.800 --> 0:34:58.800  
Amit Sethi  
OK, so the so basically when we do this, we say that just when we sum over I.

0:35:0.510 --> 0:35:1.120  
Amit Sethi  
OK, fair.

0:35:2.100 --> 0:35:11.810  
Amit Sethi  
Uh, why I the output that we get and several Pi is equal to the decision I think.

0:35:13.310 --> 0:35:14.680  
Amit Sethi  
Hmm. Excel.

0:35:19.620 --> 0:35:20.520  
Amit Sethi  
Are you such that?

0:35:22.470 --> 0:35:25.480  
Amit Sethi  
OK, so these are the correctly classic correct classification.

0:35:29.440 --> 0:35:32.910  
Amit Sethi  
OK, for this the this term.

0:35:32.960 --> 0:35:34.430  
Amit Sethi  
This multiplication will be plus one.

0:35:37.710 --> 0:35:43.100  
Amit Sethi  
OK, so here we will say that this is Wim.

0:35:45.70 --> 0:35:46.950  
Amit Sethi  
E to the power minus L5.

0:35:48.830 --> 0:35:57.950  
Amit Sethi  
And then we have the misclassified ones where the I is not equal to I can excel.

0:36:4.150 --> 0:36:9.680  
Amit Sethi  
So here we will have uh WM WIN.

0:36:13.90 --> 0:36:13.280  
Amit Sethi  
Right.

0:36:13.440 --> 0:36:16.610  
Amit Sethi  
And then we will have E to the power plus alpha.

0:36:19.530 --> 0:36:22.30  
Amit Sethi  
OK, where the signs don't agree.

0:36:23.900 --> 0:36:24.100  
Amit Sethi  
Really.

0:36:35.250 --> 0:36:41.260  
Amit Sethi  
Then we take basically all of these and put it here and then take the difference and put it in the next step.

0:36:41.270 --> 0:36:51.130  
Amit Sethi  
So basically we get sum over all I is equal to 1 to north uh WINE to the power minus L part one.

0:36:52.580 --> 0:36:52.970  
Amit Sethi  
OK.

0:36:53.440 --> 0:36:57.840  
Amit Sethi  
And then only for the misclassified ones, yeah.

0:36:59.640 --> 0:37:0.420  
Amit Sethi  
Not equal to.

0:37:2.50 --> 0:37:2.630  
Amit Sethi  
Excellent.

0:37:2.640 --> 0:37:5.450  
Amit Sethi  
PO hmm. Itson.

0:37:8.750 --> 0:37:10.640  
Amit Sethi  
And here we will have the difference.

0:37:11.90 --> 0:37:14.810  
Amit Sethi  
So we'll have a he to the power.

0:37:16.720 --> 0:37:20.890  
Amit Sethi  
Ohh it was stupid the wait.

0:37:20.940 --> 0:37:22.270  
Amit Sethi  
Yeah, we have the weight also.

0:37:29.0 --> 0:37:34.590  
Amit Sethi  
You need to be power alpha n -, E to the power minus L5.

0:37:43.840 --> 0:37:44.140  
Amit Sethi  
Alright.

0:37:57.10 --> 0:37:57.320  
Amit Sethi  
Good.

0:37:59.920 --> 0:38:10.960  
Amit Sethi  
OK, so now we have to just differentiate this and set the derivative to 0 and that will give us what is the ideal value for alpha in OK, what should be the bandwidth for the wait for the?

0:38:14.350 --> 0:38:24.600  
Amit Sethi  
So if we take and you by Dell Alpha M equal to 0 implies what we'll get is alpha is equal to.

0:38:28.830 --> 0:38:29.380  
Amit Sethi  
Off.

0:38:31.970 --> 0:38:34.310  
Amit Sethi  
Ohh, this law of missing issue.

0:38:41.590 --> 0:38:49.810  
Amit Sethi  
OK Ohh all samples divided by the ohh weight of all samples divided by weight of misclassified.

0:38:52.720 --> 0:38:53.590  
Amit Sethi  
So this is.

0:38:57.240 --> 0:39:0.30  
Amit Sethi  
Some over all line is equal to 1 to N.

0:39:3.490 --> 0:39:7.240  
Amit Sethi  
So not all this is for the idea of the base agreement.

0:39:8.880 --> 0:39:13.170  
Amit Sethi  
So wherever your TI is equal to.

0:39:14.630 --> 0:39:15.310  
Amit Sethi  
Uh.

0:39:15.740 --> 0:39:16.550  
Amit Sethi  
Hmm. XL.

0:39:21.60 --> 0:39:21.260  
Amit Sethi  
OK.

0:39:22.90 --> 0:39:23.400  
Amit Sethi  
You're gonna take care rates.

0:39:30.560 --> 0:39:35.320  
Amit Sethi  
And then divided by the misclassified sentence Pi not equal to.

0:39:38.870 --> 0:39:39.910  
Amit Sethi  
It's an excellent.

0:39:41.850 --> 0:39:42.170  
Amit Sethi  
OK.

0:39:42.470 --> 0:39:43.190  
Amit Sethi  
And they are good.

0:39:51.90 --> 0:39:51.590  
Amit Sethi  
It could.

0:39:51.630 --> 0:39:52.840  
Amit Sethi  
So what does this mean?

0:39:53.30 --> 0:39:54.150  
Amit Sethi  
When will alpha be high?

0:40:1.540 --> 0:40:7.540  
Amit Sethi  
When the last time I was there, when the numerator is high, then also terms in the numerator.

0:40:9.390 --> 0:40:12.280  
Amit Sethi  
Ohh, lots of high weight terms in the new.

0:40:14.750 --> 0:40:15.860  
Amit Sethi  
Thoughts of tones and I.

0:40:19.660 --> 0:40:23.790  
Amit Sethi  
So in the final classification you get highway to those samples.

0:40:23.800 --> 0:40:29.510  
Amit Sethi  
That so something like classify the highly rated samples but.

0:40:39.390 --> 0:40:39.710  
Amit Sethi  
I'm sorry.

0:40:43.250 --> 0:40:43.760  
Amit Sethi  
Order of the.

0:40:45.650 --> 0:40:50.90  
Amit Sethi  
Order of the model matrix, yes, each model depends on the previously trained models.

0:40:57.50 --> 0:40:57.370  
Amit Sethi  
Sorry.

0:41:3.370 --> 0:41:3.490  
Amit Sethi  
Yes.

0:41:7.720 --> 0:41:10.500  
Amit Sethi  
How is this method better over?

0:41:12.340 --> 0:41:14.740  
Amit Sethi  
Ohh or world?

0:41:14.750 --> 0:41:18.130  
Amit Sethi  
What over regular and semble, for example?

0:41:20.360 --> 0:41:37.90  
Amit Sethi  
Ohh so the way this model is better is that I think you need fewer models in the chain, whereas with ensemble you you have that law of large numbers for which you need large number of samples to be in somewhere.

0:41:38.290 --> 0:41:44.650  
Amit Sethi  
Here you can get good results with very few models because models are dependent on the previous models.

0:41:47.270 --> 0:41:49.410  
Amit Sethi  
That's the main difference I think.

0:41:51.40 --> 0:41:54.30  
Amit Sethi  
Perhaps they all converge to the same solution?

0:41:54.280 --> 0:41:54.940  
Amit Sethi  
I'm not sure.

0:41:56.320 --> 0:41:56.970  
Amit Sethi  
Ohh.

0:41:59.210 --> 0:42:1.780  
Amit Sethi  
But it gives you better results with fewer models.

0:42:8.980 --> 0:42:9.580  
Amit Sethi  
Any questions?

0:42:15.640 --> 0:42:28.210  
Amit Sethi  
So actually Bremen compared random forest with boosting also and he found that random for this actually gives better results because see here we are not doing any kind of bagging, we are not doing any kind of bootstrapping.

0:42:28.220 --> 0:42:31.830  
Amit Sethi  
We are not subsampling the samples, right?

0:42:31.880 --> 0:42:36.810  
Amit Sethi  
All samples are used in all models because only then you can keep consistent weights.

0:42:37.240 --> 0:42:40.20  
Amit Sethi  
You can consistently calculate weights from one step to the other.

0:42:42.740 --> 0:42:47.940  
Amit Sethi  
So the the disadvantage here is that you will not get out of bag estimate.

0:42:49.950 --> 0:42:51.740  
Amit Sethi  
Which you can get when you do.

0:42:51.810 --> 0:42:52.950  
Amit Sethi  
When you train random forests.

0:43:9.540 --> 0:43:11.710  
Amit Sethi  
Shall we write down the whole algorithm?

0:43:12.320 --> 0:43:13.460  
Amit Sethi  
If are you done with this?

0:43:20.930 --> 0:43:21.110  
Amit Sethi  
Yeah.

0:43:24.970 --> 0:43:25.240  
Amit Sethi  
Yep.

0:43:30.640 --> 0:43:30.850  
Amit Sethi  
Not.

0:43:32.980 --> 0:43:42.120  
Amit Sethi  
We cannot choose which model comes first because the second model only trains after the first model has trained and we know where the first model is making errors.

0:43:43.920 --> 0:43:44.340  
Amit Sethi  
So you can.

0:43:47.500 --> 0:43:49.500  
Amit Sethi  
Can you choose the order the order?

0:43:50.950 --> 0:43:56.290  
Amit Sethi  
Ohh, once they are trained they the order doesn't matter, but during training the order is everything in boosting.

0:43:57.860 --> 0:44:7.650  
Amit Sethi  
OK, so so you don't choose, you actually take the weighted average of all models anyway, but let's say you trained a chain of 100 models.

0:44:7.940 --> 0:44:13.310  
Amit Sethi  
If you're suggesting that, look at the alphas and drop the models with low alpha, perhaps you can do that.

0:44:13.660 --> 0:44:15.920  
Amit Sethi  
I don't know if someone has done the analysis of it.

0:44:16.120 --> 0:44:17.350  
Amit Sethi  
They might be papers that do that.

0:44:19.530 --> 0:44:23.940  
Amit Sethi  
So this is your entire model, so you have team models, OK.

0:44:28.360 --> 0:44:30.100  
Amit Sethi  
I think the my other slide is better.

0:44:32.780 --> 0:44:32.960  
Amit Sethi  
Yeah.

0:44:43.330 --> 0:44:43.460  
Amit Sethi  
Yeah.

0:44:48.330 --> 0:44:51.780  
Amit Sethi  
More likely to overfit if you if you train too many models with boosting.

0:44:52.740 --> 0:44:54.180  
Amit Sethi  
Yeah, it can overfit.

0:44:56.250 --> 0:45:2.960  
Amit Sethi  
So you'll have to keep a separate validation set in this case, whereas the validation set comes for free in random forest.

0:45:5.850 --> 0:45:10.230  
Amit Sethi  
So basically your your entire algorithm will be something like this, so you have.

0:45:12.610 --> 0:45:13.320  
Amit Sethi  
Misha dies.

0:45:18.230 --> 0:45:23.960  
Amit Sethi  
You're uh, WIN, uh, WI 1 right?

0:45:23.970 --> 0:45:28.300  
Amit Sethi  
For for all of these, equal to 1 by N for online.

0:45:30.80 --> 0:45:35.50  
Amit Sethi  
OK, so this is your first iteration and then.

0:45:36.810 --> 0:45:44.970  
Amit Sethi  
So this is your first initialization in the second initialization, it's your, so I guess the way it will have to calculate, right?

0:45:44.980 --> 0:45:46.990  
Amit Sethi  
So that it comes from back to that.

0:45:49.620 --> 0:45:50.50  
Amit Sethi  
Sir.

0:45:50.680 --> 0:45:51.900  
Amit Sethi  
But then for?

0:45:56.780 --> 0:46:0.830  
Amit Sethi  
For P is equal to 1, two capital K OK, these are the .K.

0:46:0.840 --> 0:46:1.630  
Amit Sethi  
Models that you have.

0:46:4.160 --> 0:46:5.270  
Amit Sethi  
We're good, then?

0:46:7.50 --> 0:46:8.330  
Amit Sethi  
Maybe not to the handout.

0:46:9.100 --> 0:46:10.40  
Amit Sethi  
Stop singing.

0:46:10.840 --> 0:46:12.420  
Amit Sethi  
OK, so.

0:46:17.120 --> 0:46:17.620  
Amit Sethi  
Learn.

0:46:20.150 --> 0:46:22.400  
Amit Sethi  
Edge uh.

0:46:24.910 --> 0:46:35.500  
Amit Sethi  
HP OK for all ohh so basically what does now mean some minimize?

0:46:41.310 --> 0:46:44.260  
Amit Sethi  
Minimize the loss, which is some loss.

0:46:44.270 --> 0:46:46.240  
Amit Sethi  
OK, some over I.

0:46:47.810 --> 0:46:48.40  
Amit Sethi  
Yeah.

0:46:50.930 --> 0:46:53.270  
Amit Sethi  
WI OK.

0:46:55.490 --> 0:46:55.940  
Amit Sethi  
For this.

0:46:58.320 --> 0:47:6.150  
Amit Sethi  
Etcherla lost lost loss over uh Edge.

0:47:7.700 --> 0:47:9.310  
Amit Sethi  
Uh, OK, I.

0:47:12.760 --> 0:47:13.600  
Amit Sethi  
And here.

0:47:16.770 --> 0:47:19.0  
Amit Sethi  
OK, so why have you small?

0:47:19.210 --> 0:47:21.500  
Amit Sethi  
Because this is the loss for her single sample.

0:47:26.230 --> 0:47:33.830  
Amit Sethi  
And usually the loss for all the samples stumble word individual losses, but here we will take weighted weighted sum.

0:47:36.340 --> 0:47:38.960  
Amit Sethi  
Minimize how will you minimize over parameters.

0:47:45.30 --> 0:47:46.490  
Amit Sethi  
Off what?

0:47:46.910 --> 0:47:50.590  
Amit Sethi  
HQ, whatever is a you've learned or HQ, whatever is your model.

0:47:52.540 --> 0:47:56.40  
Amit Sethi  
Pick we minimize over the parameters of worksheet.

0:47:56.540 --> 0:47:57.500  
Amit Sethi  
This particular loss.

0:47:59.310 --> 0:47:59.630  
Amit Sethi  
Alright.

0:48:6.400 --> 0:48:7.430  
Amit Sethi  
Then your error rate.

0:48:10.470 --> 0:48:13.430  
Amit Sethi  
Epsilon N Epsilon thing for the teeth model.

0:48:15.240 --> 0:48:15.750  
Amit Sethi  
Is the.

0:48:19.170 --> 0:48:20.730  
Amit Sethi  
Basically the sum of.

0:48:23.470 --> 0:48:25.240  
Amit Sethi  
Ohh the waiting error.

0:48:27.290 --> 0:48:31.250  
Amit Sethi  
So this will be ohh some or what I.

0:48:33.950 --> 0:48:34.910  
Amit Sethi  
WIK.

0:48:39.760 --> 0:48:43.850  
Amit Sethi  
And the indicator function where UH-2 disagree.

0:48:46.110 --> 0:48:49.570  
Amit Sethi  
HI is not equal to.

0:48:55.390 --> 0:48:56.230  
Amit Sethi  
So basically your error.

0:48:59.910 --> 0:49:2.140  
Amit Sethi  
20% point two, etcetera.

0:49:3.150 --> 0:49:3.330  
Amit Sethi  
OK.

0:49:4.140 --> 0:49:5.600  
Amit Sethi  
And then your video.

0:49:7.650 --> 0:49:9.960  
Amit Sethi  
This is assignment we.

0:49:17.980 --> 0:49:19.530  
Amit Sethi  
It was not free Alpha King.

0:49:21.980 --> 0:49:25.540  
Amit Sethi  
My notations are messed up now, but keep that.

0:49:26.310 --> 0:49:27.270  
Amit Sethi  
So what is alpha?

0:49:27.280 --> 0:49:29.790  
Amit Sethi  
OK, this will be half.

0:49:30.980 --> 0:49:31.760  
Amit Sethi  
Remember that dog?

0:49:31.770 --> 0:49:40.780  
Amit Sethi  
That really calculated right, correct or incorrect, so this will be 1 minus epsilon K divided by epsilon.

0:49:40.790 --> 0:49:44.720  
Amit Sethi  
OK, so basically it will be, uh, one way excellent game.

0:49:47.70 --> 0:49:47.530  
Amit Sethi  
Minus one.

0:49:49.430 --> 0:49:50.950  
Amit Sethi  
This is the read for the model.

0:49:57.870 --> 0:50:4.530  
Amit Sethi  
And then we need to find wait for the samples for the K + 1 iteration.

0:50:9.910 --> 0:50:10.140  
Amit Sethi  
OK.

0:50:12.580 --> 0:50:13.630  
Amit Sethi  
And what was that?

0:50:13.680 --> 0:50:14.350  
Amit Sethi  
That was the.

0:50:16.690 --> 0:50:17.230  
Amit Sethi  
Ohh.

0:50:19.300 --> 0:50:26.0  
Amit Sethi  
Number we had that, uh, so great that we had after.

0:50:28.910 --> 0:50:35.610  
Amit Sethi  
OK, uh, times in your speed or Italy power?

0:50:38.230 --> 0:50:38.550  
Amit Sethi  
What?

0:50:38.990 --> 0:50:39.880  
Amit Sethi  
Ohh my.

0:50:41.750 --> 0:50:49.290  
Amit Sethi  
Minus WK, what is the I changed WI this alpha.

0:50:53.550 --> 0:50:54.850  
Amit Sethi  
Sorry for the mess.

0:50:55.160 --> 0:50:55.500  
Amit Sethi  
Alright.

0:50:55.900 --> 0:51:1.520  
Amit Sethi  
And then we will have EI and the whole basic can I?

0:51:3.410 --> 0:51:5.510  
Amit Sethi  
TI times by.

0:51:7.700 --> 0:51:9.140  
Amit Sethi  
By I at the M step.

0:51:11.810 --> 0:51:14.590  
Amit Sethi  
Divided by some the same thing as a sum.

0:51:18.570 --> 0:51:20.20  
Amit Sethi  
Is equal to 1 to 10.

0:51:22.140 --> 0:51:25.900  
Amit Sethi  
You should be power minus Lok for.

0:51:28.160 --> 0:51:33.500  
Amit Sethi  
Believe JYJ at the end of the day.

0:51:36.90 --> 0:51:38.50  
Amit Sethi  
Then why is the complete decision?

0:51:38.870 --> 0:51:39.220  
Amit Sethi  
Yes.

0:51:39.360 --> 0:51:49.130  
Amit Sethi  
So the complete decision tells you though wait that you will have so here if you correctly classify then the wait for that sample will reduce.

0:51:50.700 --> 0:52:1.760  
Amit Sethi  
OK, if you incorrectly laughing then the wait will increase because your denominator will have that the product of T&Y will be negative and when the negative in the exponent it will become positive.

0:52:2.320 --> 0:52:9.690  
Amit Sethi  
So the way it will increase and that way it will make sure that the next model gives it more importance.

0:52:17.690 --> 0:52:17.910  
Amit Sethi  
OK.

0:52:19.860 --> 0:52:21.490  
Amit Sethi  
So that's your complete thing.

0:52:21.600 --> 0:52:28.70  
Amit Sethi  
Your complete decision making and basically your model Yi.

0:52:30.460 --> 0:52:32.810  
Amit Sethi  
Why are you at the Amit step?

0:52:32.880 --> 0:52:33.400  
Amit Sethi  
Any amount?

0:52:34.610 --> 0:52:37.810  
Amit Sethi  
It should be clear from using thing.

0:52:44.730 --> 0:52:47.320  
Amit Sethi  
Then what happens when you try to reach from too many sources?

0:52:51.40 --> 0:52:51.290  
Amit Sethi  
Why?

0:52:51.300 --> 0:52:54.170  
Amit Sethi  
I had any step is sign.

0:52:56.160 --> 0:52:57.970  
Amit Sethi  
OK, of the.

0:53:1.820 --> 0:53:3.970  
Amit Sethi  
Of all the hypothesis that you had earlier.

0:53:5.860 --> 0:53:9.350  
Amit Sethi  
So this will be WI.

0:53:12.620 --> 0:53:15.940  
Amit Sethi  
If you I I want to use M now because here I will have some.

0:53:18.160 --> 0:53:18.900  
Amit Sethi  
Some of.

0:53:25.740 --> 0:53:28.660  
Amit Sethi  
Some of every is equal to 1 to K.

0:53:34.310 --> 0:53:36.130  
Amit Sethi  
And you will have whatever that.

0:53:38.970 --> 0:53:41.330  
Amit Sethi  
XN it's like.

0:53:47.890 --> 0:53:49.270  
Amit Sethi  
And we'll also have alpha.

0:53:55.840 --> 0:53:59.30  
Amit Sethi  
Our space let me let me write this second.

0:53:59.40 --> 0:54:0.50  
Amit Sethi  
So there's cleaner.

0:54:3.990 --> 0:54:6.520  
Amit Sethi  
So basically it will be some of all the models right?

0:54:6.530 --> 0:54:7.400  
Amit Sethi  
So it will be.

0:54:7.990 --> 0:54:10.540  
Amit Sethi  
You will take sign whether it is positive or negative.

0:54:10.650 --> 0:54:14.590  
Amit Sethi  
Then you will take some is sum over what sum over all the models.

0:54:14.690 --> 0:54:20.380  
Amit Sethi  
OK, so M is equal to 1 2K here up to the other one.

0:54:20.520 --> 0:54:24.790  
Amit Sethi  
OK, so you will have a the.

0:54:26.820 --> 0:54:27.680  
Amit Sethi  
Uh something?

0:54:27.690 --> 0:54:39.140  
Amit Sethi  
Yeah, you will have the model weight which is alpha M OK and then you'll have the decision and then the decision will be H of HMM Auto Excel.

0:54:40.230 --> 0:54:40.700  
Amit Sethi  
This is it.

0:54:42.830 --> 0:54:44.620  
Amit Sethi  
So this is your decision at anything.

0:55:7.20 --> 0:55:7.730  
Amit Sethi  
Any question?

0:55:17.610 --> 0:55:17.870  
Amit Sethi  
Oil.

0:55:17.950 --> 0:55:25.60  
Amit Sethi  
Some books so that the weights add up to one the same expression in the numerator, but for all sample.

0:55:27.720 --> 0:55:28.270  
Amit Sethi  
This one right?

0:55:30.720 --> 0:55:32.150  
Amit Sethi  
This is all samples.

0:55:33.420 --> 0:55:35.630  
Amit Sethi  
Ohh so that all may set up to.

0:55:35.680 --> 0:55:37.890  
Amit Sethi  
Initially they are all one by one, so they add up to one.

0:55:50.140 --> 0:55:51.60  
Amit Sethi  
Ohh.

0:55:54.780 --> 0:55:55.70  
Amit Sethi  
Yeah.

0:55:55.80 --> 0:55:55.390  
Amit Sethi  
Yeah.

0:55:55.400 --> 0:55:57.100  
Amit Sethi  
So here you have the UI so.

0:55:59.220 --> 0:55:59.670  
Amit Sethi  
But we do.

0:56:0.960 --> 0:56:4.810  
Amit Sethi  
Yeah, exactly the same expression as the numerator.

0:56:5.40 --> 0:56:5.640  
Amit Sethi  
Yeah. Thanks.

0:56:27.440 --> 0:56:28.890  
Amit Sethi  
So it's a Mehta procedure.

0:56:28.900 --> 0:56:36.300  
Amit Sethi  
It doesn't care about what the weak learner is, as long as the loss can be summed over and it can, it can compute the weighted sum.

0:56:38.830 --> 0:56:44.720  
Amit Sethi  
So for example, if you're doing a cross entropy right, you could have each.

0:56:44.890 --> 0:56:47.760  
Amit Sethi  
Each model could be a linear regress linear classifier.

0:56:48.370 --> 0:56:51.120  
Amit Sethi  
So let's say if you're crossing topic will be weighted crossing.

0:56:51.730 --> 0:56:59.510  
Amit Sethi  
So here your loss that you're minimizing will become weighted cross entropy weighted by the weight of each sample.

0:57:2.440 --> 0:57:11.60  
Amit Sethi  
When you actually look for weighted cross entropy in scandal or something, then they're the wait is for each class.

0:57:12.380 --> 0:57:16.350  
Amit Sethi  
That's a different your greatest, but not for each class.

0:57:17.160 --> 0:57:20.520  
Amit Sethi  
We have to use classifier morning.

0:57:21.310 --> 0:57:21.890  
Amit Sethi  
It's no, no.

0:57:23.820 --> 0:57:28.280  
Amit Sethi  
So it will be 1 Model, 1 media classifier, another linear classifier.

0:57:28.290 --> 0:57:36.280  
Amit Sethi  
And remember, initially I showed you that if you have multiple linear classifiers, you if you simply take their average you can get a nonlinear classifier.

0:57:37.330 --> 0:57:38.340  
Amit Sethi  
So same concept here.

0:57:38.350 --> 0:57:41.620  
Amit Sethi  
Instead of average, it's weighted and weighted by alpha.

0:57:44.950 --> 0:57:46.340  
Amit Sethi  
And how do we get different alphas?

0:57:46.350 --> 0:57:59.220  
Amit Sethi  
Because we can give different weights submits to each sample, how do you get get different rates to each sample by looking at how much error has been made on this sample in the in all M -, 1, classifieds?

0:58:1.220 --> 0:58:9.20  
Amit Sethi  
And how do we get be classified based on how I can then activate like dollar of numerator very moment.

0:58:9.90 --> 0:58:10.320  
Amit Sethi  
That the person either.

0:58:18.30 --> 0:58:18.960  
Amit Sethi  
You can use any function.

0:58:20.880 --> 0:58:22.260  
Amit Sethi  
Because you you wanna Make Love?

0:58:22.270 --> 0:58:22.700  
Amit Sethi  
I don't.

0:58:23.190 --> 0:58:23.760  
Amit Sethi  
I will give.

0:58:24.510 --> 0:58:30.0  
Amit Sethi  
I have that that that talk that into this thing.

0:58:30.50 --> 0:58:36.840  
Amit Sethi  
Ohh get easier once because you are changing the standard one based on there and you.

0:58:38.830 --> 0:58:40.570  
Amit Sethi  
On this, I believe you.

0:58:40.680 --> 0:58:41.170  
Amit Sethi  
You know I haven't.

0:58:41.490 --> 0:58:44.760  
Amit Sethi  
OK, you have so that that.

0:58:46.860 --> 0:58:52.130  
Amit Sethi  
But you we also depending what happens, right, right.

0:58:52.370 --> 0:58:56.310  
Amit Sethi  
And you know, so that that's why we don't listen to that.

0:58:57.480 --> 0:59:2.280  
Amit Sethi  
So you can for example use, let's say each of these send your threshold logic.

0:59:4.40 --> 0:59:8.30  
Amit Sethi  
Now if you had any, the Lords already decided, right?

0:59:8.40 --> 0:59:9.240  
Amit Sethi  
But that's not fair.

0:59:9.260 --> 0:59:10.170  
Amit Sethi  
So you can use that.

0:59:11.500 --> 0:59:13.180  
Amit Sethi  
Just that you will evaluate this loss.

0:59:14.640 --> 0:59:29.60  
Amit Sethi  
OK, you can just evaluate this loss with the weight with a bit and you will evaluate it for all the choices of are you going to send next Harshit that.

0:59:29.830 --> 0:59:30.170  
Amit Sethi  
Thank you.

0:59:31.170 --> 0:59:36.680  
Amit Sethi  
Just so you evaluated talking for all of them, you said you will evaluate this loss again.

0:59:36.770 --> 0:59:38.210  
Amit Sethi  
Let's see who's the background music.

0:59:40.230 --> 0:59:40.640  
Amit Sethi  
Let's see.

0:59:40.910 --> 0:59:41.550  
Amit Sethi  
I think yeah.

0:59:42.470 --> 0:59:45.960  
Amit Sethi  
Sometime this morning or sometime this morning?

0:59:46.410 --> 0:59:46.720  
Amit Sethi  
No.

0:59:46.730 --> 0:59:58.700  
Amit Sethi  
So here you cannot some sample the samples and he's like you cannot is because you're computing a wait for all samples give when you're computing a wait for all samples.

0:59:59.110 --> 1:0:5.270  
Amit Sethi  
That only works because the the the nominator is like the denominator over all sentence.

1:0:6.120 --> 1:0:8.360  
Amit Sethi  
So I don't know if you can.

1:0:10.320 --> 1:0:14.50  
Amit Sethi  
Maybe you can adapt it for subsampling also, but the original algorithm.

1:0:16.900 --> 1:0:21.510  
Amit Sethi  
You or you have to assume that the training set remains the same across multiple mode.

1:0:22.470 --> 1:0:23.250  
Amit Sethi  
I'm like kind of.

1:0:30.510 --> 1:0:30.730  
Amit Sethi  
Yep.

1:0:59.230 --> 1:1:1.50  
Amit Sethi  
OK. Ohh.

1:1:3.660 --> 1:1:4.210  
Amit Sethi  
And it is this.

1:1:15.670 --> 1:1:19.290  
Amit Sethi  
So now we'll look at another kind of boosting, which is the gradient listing.

1:1:29.30 --> 1:1:38.150  
Amit Sethi  
Uh, But here the setup is that again your next model, you have it, it's actually original somewhere.

1:1:38.160 --> 1:1:39.330  
Amit Sethi  
Formulation is for regression.

1:1:39.600 --> 1:1:45.270  
Amit Sethi  
In gradient boosting it's not for, it's not for classification, but people have adapted it to classification.

1:1:45.280 --> 1:1:47.460  
Amit Sethi  
So that's why instead of C I'm using F.

1:1:50.850 --> 1:2:6.630  
Amit Sethi  
But is, uh, if you take it over any sample excel, it is uh, the previous model up to a month set and last we'll have the new model.

1:2:13.620 --> 1:2:18.40  
Amit Sethi  
And so this is basically what you what you have alright.

1:2:26.250 --> 1:2:26.660  
Amit Sethi  
OK.

1:2:27.50 --> 1:2:31.0  
Amit Sethi  
So, uh, the laws that you will have.

1:2:31.980 --> 1:2:33.140  
Amit Sethi  
Good, no.

1:2:34.900 --> 1:2:36.570  
Amit Sethi  
So we want to approximate this.

1:2:39.440 --> 1:2:42.670  
Amit Sethi  
This uh, adding of the next model, right?

1:2:42.900 --> 1:2:44.500  
Amit Sethi  
So your emeth model.

1:2:53.870 --> 1:2:57.390  
Amit Sethi  
Ohh will be the previous model.

1:3:3.220 --> 1:3:11.150  
Amit Sethi  
And or do we try to do with this is you will try it with the next model that you are adding.

1:3:11.160 --> 1:3:13.600  
Amit Sethi  
Hmm, you will try to do.

1:3:15.620 --> 1:3:16.500  
Amit Sethi  
Reduce the residue.

1:3:17.870 --> 1:3:25.940  
Amit Sethi  
Everyone knows what's the residual but error so here if it is, let's say it is meant for ohm.

1:3:26.860 --> 1:3:29.330  
Amit Sethi  
This is meant for regression.

1:3:29.580 --> 1:3:39.190  
Amit Sethi  
So basically your TI OK, minus whatever is the previous model like the buy whatever is your buy it right.

1:3:39.200 --> 1:3:40.170  
Amit Sethi  
So that is your best signal.

1:3:41.320 --> 1:3:44.250  
Amit Sethi  
OK, so we want to reduce the residual.

1:3:44.260 --> 1:3:45.180  
Amit Sethi  
How do you reduce it?

1:3:45.190 --> 1:3:47.710  
Amit Sethi  
You basically take a step in the negative side of the gradient.

1:3:48.940 --> 1:3:49.260  
Amit Sethi  
We're good.

1:3:50.420 --> 1:3:59.620  
Amit Sethi  
So this will be minus some stepsons OK and then some forward all the samples.

1:4:3.240 --> 1:4:3.690  
Amit Sethi  
OK.

1:4:3.940 --> 1:4:8.500  
Amit Sethi  
And then you will take the gradient off FM minus one.

1:4:12.270 --> 1:4:12.540  
Amit Sethi  
Uh.

1:4:15.780 --> 1:4:20.860  
Amit Sethi  
Agreeing with the respect to the parameters of if and then this one.

1:4:24.70 --> 1:4:24.500  
Amit Sethi  
OK.

1:4:24.770 --> 1:4:25.880  
Amit Sethi  
And some loss.

1:4:28.310 --> 1:4:30.780  
Amit Sethi  
And this loss will be fit on this TI.

1:4:32.820 --> 1:4:37.920  
Amit Sethi  
And the error that you had not sent to the space.

1:4:40.620 --> 1:4:42.960  
Amit Sethi  
F n -, 1 XL.

1:4:47.280 --> 1:4:49.610  
Amit Sethi  
OK, So what is this loss?

1:4:50.100 --> 1:4:52.590  
Amit Sethi  
This loss can be something like mean square error.

1:5:2.250 --> 1:5:3.190  
Amit Sethi  
But unlike.

1:5:4.770 --> 1:5:7.340  
Amit Sethi  
Ohh, can you hear turning there?

1:5:7.350 --> 1:5:10.550  
Amit Sethi  
We keep the learning rate to be small here.

1:5:10.560 --> 1:5:14.40  
Amit Sethi  
This drama parameter that we have, it will not be stopped.

1:5:14.370 --> 1:5:19.500  
Amit Sethi  
We will actually search what is gamma that actually minimizes logs.

1:5:20.830 --> 1:5:22.260  
Amit Sethi  
OK, that will minimize this.

1:5:22.750 --> 1:5:25.870  
Amit Sethi  
So your gamma M they just have been done.

1:5:25.880 --> 1:5:30.790  
Amit Sethi  
OK, so the Aman M will be RN.

1:5:33.280 --> 1:5:34.250  
Amit Sethi  
Using a line search.

1:5:35.650 --> 1:5:36.80  
Amit Sethi  
OK.

1:5:36.130 --> 1:5:37.80  
Amit Sethi  
Are we know what number?

1:5:39.990 --> 1:5:42.760  
Amit Sethi  
Ohh, that minimizes this loss.

1:5:45.650 --> 1:5:47.760  
Amit Sethi  
Some of invisible to 1:00 PM.

1:5:50.800 --> 1:5:53.90  
Amit Sethi  
Loss of Pi.

1:5:55.860 --> 1:5:58.610  
Amit Sethi  
And now this FM.

1:6:16.250 --> 1:6:16.580  
Amit Sethi  
You could.

1:6:25.170 --> 1:6:28.930  
Amit Sethi  
OK, so every time what we are doing is we are ohh.

1:6:31.130 --> 1:6:35.420  
Amit Sethi  
Basically this becomes when you look at this loss like MSC, right?

1:6:35.730 --> 1:6:38.820  
Amit Sethi  
So if you have P, -, Y, right.

1:6:38.830 --> 1:6:41.880  
Amit Sethi  
So if you take the square, what is the gradient of this?

1:6:41.890 --> 1:6:49.190  
Amit Sethi  
It is 2 times with respect to why I will be minus 2 \* P I mean this one.

1:6:51.180 --> 1:6:54.180  
Amit Sethi  
The that will be the gradient with respect to buyer.

1:6:54.990 --> 1:7:4.880  
Amit Sethi  
So if this is the residual, it becomes the the gradient becomes my for minus two times the residual OK.

1:7:6.800 --> 1:7:10.960  
Amit Sethi  
Uh, so basically what you are doing is you are trying to.

1:7:13.730 --> 1:7:18.750  
Amit Sethi  
Take the residual and you are basically fitting a model that will go towards the.

1:7:19.70 --> 1:7:21.40  
Amit Sethi  
Try to reduce that risk every time.

1:7:21.230 --> 1:7:25.390  
Amit Sethi  
So every time you will compute the residual and go towards the negative side of that residual.

1:7:26.920 --> 1:7:30.490  
Amit Sethi  
OK, so this becomes your complete model.

1:7:35.460 --> 1:7:36.560  
Amit Sethi  
Your complete algorithm.

1:7:41.640 --> 1:7:41.990  
Amit Sethi  
Right.

1:7:42.400 --> 1:7:47.270  
Amit Sethi  
So you have from loop from M is equal to 1.

1:7:51.660 --> 1:7:53.800  
Amit Sethi  
This is your move for adding models.

1:7:56.60 --> 1:8:4.370  
Amit Sethi  
And in this basically you will first compute the residuals RIN which is the error that you have at the Amit stage.

1:8:4.920 --> 1:8:5.270  
Amit Sethi  
OK.

1:8:5.690 --> 1:8:24.810  
Amit Sethi  
And that will be equal to negative of the gradient, so negative of L boss OK, no API from if Excel whatever we have until now.

1:8:27.940 --> 1:8:29.770  
Amit Sethi  
OK by 10.

1:8:31.70 --> 1:8:33.850  
Amit Sethi  
Ohh, the parameters of right.

1:8:44.0 --> 1:8:45.450  
Amit Sethi  
We had to go X.

1:8:48.230 --> 1:8:48.970  
Amit Sethi  
I think so.

1:8:49.760 --> 1:8:51.710  
Amit Sethi  
Is what that is the previous model.

1:8:52.880 --> 1:8:56.140  
Amit Sethi  
If in minus one, exit.

1:9:4.460 --> 1:9:6.790  
Amit Sethi  
OK, then this is your first step.

1:9:6.830 --> 1:9:7.340  
Amit Sethi  
Compute this.

1:9:9.380 --> 1:9:12.240  
Amit Sethi  
Your second step is fit a model.

1:9:13.640 --> 1:9:14.640  
Amit Sethi  
Pick a new model.

1:9:19.520 --> 1:9:22.510  
Amit Sethi  
New base learner that will. Ohh.

1:9:24.880 --> 1:9:27.460  
Amit Sethi  
But that will basically try to reduce the.

1:9:29.940 --> 1:9:31.390  
Amit Sethi  
No, the recipe.

1:9:34.420 --> 1:9:35.330  
Amit Sethi  
Or we use that as.

1:9:46.590 --> 1:9:50.140  
Amit Sethi  
So this will be the H and excel.

1:9:55.540 --> 1:10:0.550  
Amit Sethi  
HX, even and 3rd would be.

1:10:2.380 --> 1:10:2.990  
Amit Sethi  
Compute.

1:10:5.510 --> 1:10:5.730  
Amit Sethi  
OK.

1:10:8.420 --> 1:10:11.210  
Amit Sethi  
Compute dama him using line search.

1:10:11.260 --> 1:10:15.750  
Amit Sethi  
That will minimize the overall ohh loss of the order.

1:10:24.260 --> 1:10:25.230  
Amit Sethi  
That minimizes.

1:10:29.410 --> 1:10:33.970  
Amit Sethi  
Or one of those which is so some work.

1:10:35.270 --> 1:10:36.440  
Amit Sethi  
I really don't want to go in.

1:10:38.0 --> 1:10:38.390  
Amit Sethi  
OK.

1:10:38.680 --> 1:10:44.990  
Amit Sethi  
The entire model, which is loss over for TI while.

1:10:46.550 --> 1:10:48.710  
Amit Sethi  
Ohh FM minus one.

1:10:54.280 --> 1:10:58.100  
Amit Sethi  
Plus, the wait times when you know.

1:11:11.450 --> 1:11:13.320  
Amit Sethi  
OK, this is your room of it.

1:11:18.420 --> 1:11:21.780  
Amit Sethi  
A new new like a composite mode.

1:11:30.860 --> 1:11:31.230  
Amit Sethi  
OK.

1:11:31.280 --> 1:11:33.210  
Amit Sethi  
And this is your new individual.

1:11:43.780 --> 1:11:51.330  
Amit Sethi  
Edge is your individual model that the whole thing is your combined FM and then lasses.

1:11:51.660 --> 1:11:56.920  
Amit Sethi  
Ohh update update which is the expression given.

1:11:56.930 --> 1:12:0.620  
Amit Sethi  
Hear anything from test M -?

1:12:0.630 --> 1:12:1.930  
Amit Sethi  
One updated.

1:12:8.850 --> 1:12:9.110  
Amit Sethi  
OK.

1:12:15.440 --> 1:12:25.110  
Amit Sethi  
Moles under scaling HFM as in so basically first you find your HMM based on.

1:12:26.90 --> 1:12:28.10  
Amit Sethi  
So hmm, is your weak learner, right?

1:12:28.220 --> 1:12:32.920  
Amit Sethi  
So you will add hmm, you will change the parameters of HM such that you minimize this loss.

1:12:36.550 --> 1:12:37.770  
Amit Sethi  
So first you find.

1:12:40.80 --> 1:12:43.660  
Amit Sethi  
So your parameters of hmm will be coming from the gradient.

1:12:44.710 --> 1:12:45.80  
Amit Sethi  
OK.

1:12:45.90 --> 1:12:48.290  
Amit Sethi  
Then you will be scaling the the entire gradient using gamma.

1:12:55.790 --> 1:12:56.200  
Amit Sethi  
Yeah.

1:12:56.350 --> 1:12:56.700  
Amit Sethi  
So.

1:12:56.710 --> 1:12:57.540  
Amit Sethi  
So think of it as.

1:12:57.550 --> 1:12:59.340  
Amit Sethi  
Let's say you have of.

1:12:59.530 --> 1:13:6.580  
Amit Sethi  
You have any other connection, so the weights where does where the weights of hmm come from?

1:13:7.110 --> 1:13:10.20  
Amit Sethi  
The way it's simply come from the gradient itself.

1:13:12.310 --> 1:13:17.710  
Amit Sethi  
You're which is the residual, the variant of the residual radiate of the list.

1:13:18.400 --> 1:13:19.570  
Amit Sethi  
It's just something from that.

1:13:19.820 --> 1:13:21.350  
Amit Sethi  
But then the weights can be seen.

1:13:22.560 --> 1:13:28.320  
Amit Sethi  
All right, can be scaled together using a gun so that you just find using lineset.

1:13:36.100 --> 1:13:36.330  
Amit Sethi  
It's.

1:13:39.950 --> 1:13:40.980  
Amit Sethi  
And with the new one.

1:13:41.390 --> 1:13:43.260  
Amit Sethi  
So let's say it's a linear linear regression.

1:13:44.280 --> 1:13:46.150  
Amit Sethi  
Critical England what does the change have?

1:13:46.160 --> 1:13:49.150  
Amit Sethi  
It has all the ohh the biggest.

1:13:58.880 --> 1:13:59.70  
Amit Sethi  
It's.

1:14:1.850 --> 1:14:2.850  
Amit Sethi  
Ohh.

1:14:7.120 --> 1:14:7.590  
Amit Sethi  
Yeah.

1:14:7.600 --> 1:14:9.640  
Amit Sethi  
So here you will when you take the gradient.

1:14:11.830 --> 1:14:12.490  
Amit Sethi  
There's a reason.

1:14:21.180 --> 1:14:21.970  
Amit Sethi  
Uh, yeah.

1:14:21.980 --> 1:14:22.550  
Amit Sethi  
So when you.

1:14:23.960 --> 1:14:27.370  
Amit Sethi  
Ohh yeah, so basically with the sum over all the idea.

1:14:29.240 --> 1:14:30.720  
Amit Sethi  
Right, otherwise it is the.

1:14:30.790 --> 1:14:33.730  
Amit Sethi  
It is for only individual Yep.

1:14:41.470 --> 1:14:45.890  
Amit Sethi  
You are scaling the output of the model, but by extension you are scaling the weights of the model also.

1:15:4.640 --> 1:15:6.490  
Amit Sethi  
OK, so this is gradient boosting.

1:15:6.540 --> 1:15:11.580  
Amit Sethi  
The HG Boost is extreme gradient boosting that only does.

1:15:12.480 --> 1:15:17.570  
Amit Sethi  
Ohh it not only does the first derivative, but it also takes the headship it takes the second derivative.

1:15:17.610 --> 1:15:20.320  
Amit Sethi  
So sorry, Newton Raphson extension of this method.

1:15:23.70 --> 1:15:28.860  
Amit Sethi  
So the famous algorithm exhibits is basically gradient boosting, but with not just gradient.

1:15:28.910 --> 1:15:29.610  
Amit Sethi  
Also that action.

1:15:32.570 --> 1:15:37.120  
Amit Sethi  
And it uses like short trees.

1:15:37.130 --> 1:15:43.540  
Amit Sethi  
It trains short tree or the used to be moments, and that seems to be very good results with a few setups.

1:15:46.590 --> 1:16:2.620  
Amit Sethi  
To exchange boost random forest in the like good choices to have when you are fitting on tabular data, especially when it is when you don't care about whether the data is discrete or continuous or a mixture, then this method will go quick.

1:16:18.60 --> 1:16:19.110  
Amit Sethi  
Any last questions?

1:16:19.120 --> 1:16:20.120  
Amit Sethi  
You kind of run out of time?

1:16:27.370 --> 1:16:42.940  
Amit Sethi  
And so I mean from you should know what is gradient boosting, but sure I think if you understand and of those very well then I think that's a good place to be call try to read up on the gradient boosting as well.

1:16:42.950 --> 1:16:45.880  
Amit Sethi  
But uh, up to add up, add a boost.

1:16:46.170 --> 1:16:46.910  
Amit Sethi  
You should definitely.

1:16:49.650 --> 1:16:57.900  
Amit Sethi  
The complete derivation with the exponential loss and how we get to the ohh individual waits for samples and waits for the for the moments.

1:17:3.840 --> 1:17:4.460  
Amit Sethi  
Oh, OK.

1:17:5.970 --> 1:17:6.930  
Amit Sethi  
Any questions?

1:17:7.150 --> 1:17:8.410  
Amit Sethi  
Otherwise, I have some announcements.

1:17:11.950 --> 1:17:20.630  
Amit Sethi  
Uh, so announcement that I have not yet updated your assignment one and oh wait some months I will update it by this weekend on Google.

1:17:26.660 --> 1:17:27.940  
Amit Sethi  
Are you done with assignment too?

1:17:30.450 --> 1:17:30.800  
Amit Sethi  
Was easy.

1:17:33.430 --> 1:17:33.890  
Amit Sethi  
Lending.

1:17:35.340 --> 1:17:41.0  
Amit Sethi  
So what I'll do is I think I'll increase the weight of the segments and assignment.

1:17:42.290 --> 1:17:43.430  
Amit Sethi  
There will be no assignment.

1:17:43.770 --> 1:17:51.890  
Amit Sethi  
Or do you want to do something on visualization and and but last year?

1:17:54.170 --> 1:17:56.80  
Amit Sethi  
So what I'll do is I'll make that optional.

1:17:56.130 --> 1:17:59.0  
Amit Sethi  
So those who cannot think of a good project, they can do an assignment.

1:18:3.40 --> 1:18:8.300  
Amit Sethi  
And what I will also do is I'll post some reports of good projects over this weekend.

1:18:10.470 --> 1:18:17.610  
Amit Sethi  
So do you think of a good project I'll in the next class when you come for NSM and the transformer?

1:18:18.530 --> 1:18:18.870  
Amit Sethi  
Uh.

1:18:19.50 --> 1:18:31.220  
Amit Sethi  
You can also ask that teenagers about whether your project proposal is good or not, or you can also second time for the online consultation with the teams.

1:18:34.190 --> 1:18:35.700  
Amit Sethi  
Or just send me messages.

1:18:35.710 --> 1:18:36.540  
Amit Sethi  
I'll also respond.

1:18:38.680 --> 1:18:47.660  
Amit Sethi  
But try to do it sooner than later next week, Tuesday to Sunday, I will be traveling, so I may not have access to to teams, so try to do it before Tuesday.

1:18:58.110 --> 1:18:59.810  
Amit Sethi  
Regarding the projection, yeah.

1:18:59.860 --> 1:19:5.310  
Amit Sethi  
So the assignment three which you will give whenever equal weight is right with the project or like.