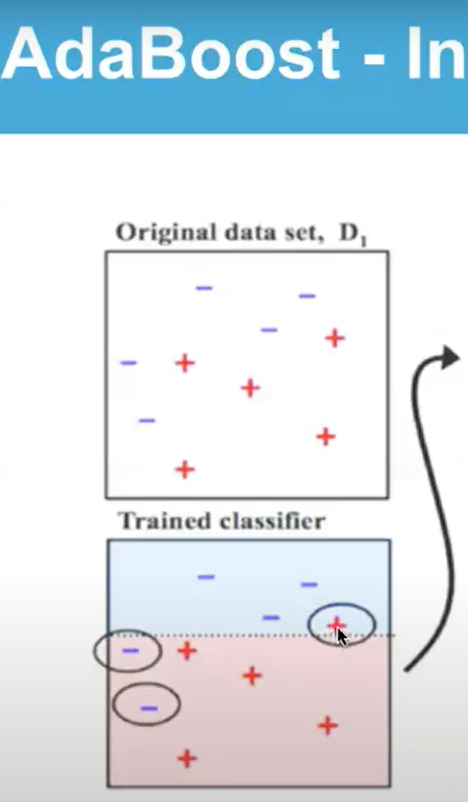
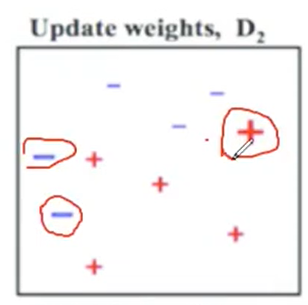
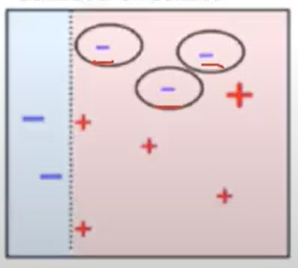
Boosting

Friday, January 1, 2021

7:53 PM

* Does really well with weak learners (unlike decision trees and random forests doing well with strong learners)
* Try to correct answer of predecessor
* Adaboost - adaptive boosting
* Bagging - all estimators are built in parallel from a subset of the sample, and all samples are weighed equally
* Boosting - some samples bigger than other, built one after the other, weights of observations adjusted in sequence and new estimator made
* Adaboost - adaptive in terms of the weight of the observations and estimators, unequal weights for both estimators and observations
* Boosting - statistical framework - objective to minimize the loss of the model by adding weak learners using a gradient descent like procedure

* Weak learner with high bias
* 
* The ones circled are the misclassified cases
* A simple model has more bias - one question is being asked (is it above or below this threshold)
* Increase weights of trouble makers - more likely to be over represent in the training stage
* 
* New circles are new trouble makers
* 
* At the end the three classifiers are combined into a strong learner, the better the classifier the more say it has (the worse the classifier the less say it has)

* All the weight has to add up to 1 or 100%
* At the onset each observation has equal weight
* Alpha is the amount of say each classifier has
* I - indicator function, I takes on one if wrongly classified
* I - if it y1 is equal to FM(xi) then I is -1
* Misclassified scales up and correctly classified scale down
* G(x) - amount of say at the final decision

* Goal of gradient boosting is to predict the residual