

Boosting

Cynthia Rudin

Machine Learning Course, Duke

Boosting Motivation

- Question of Kearns: Can you turn a “weak” learning algorithm (that is barely better than random guessing) into a “strong” learning algorithm (whose error rate is arbitrarily close to 0)?
- We could ask the algorithm to create a lot of classifiers and figure out how to combine them... how to do that?

Boosting Motivation

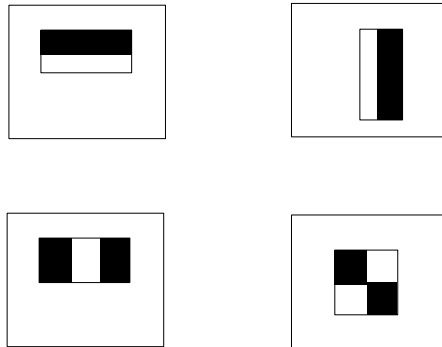
Schapire and Freund's answer:

- Reweight the data in many ways
- Use the weak learning algorithm to create a weak classifier for each (reweighted) dataset
- Compute a weighted average of the weak classifiers.

Weak classifiers used by Viola and Jones

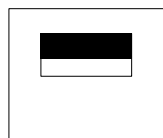
Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones

Doesn't detect
anything



Black and white areas
are very similar

Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones

Doesn't detect anything



Black and white areas are very similar

Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones

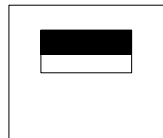
Now it detects!



Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones

Weak classifiers

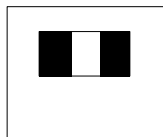
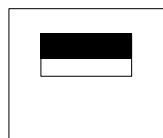


Detects eyes!

Weak classifiers used by Viola and Jones

- Subtract the white areas from the black ones

Weak classifiers



Detects eyes!



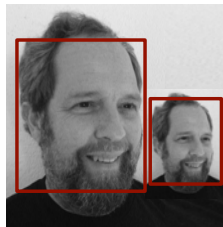
Weak classifiers used by Viola and Jones

- Used hundreds of thousands of these weak classifiers at all different scales



Weak classifiers used by Viola and Jones

- Used hundreds of thousands of these weak classifiers at all different scales



AdaBoost Pseudocode

Assign observation i the weight of $d_{1i}=1/n$ (equal weights).

For $t=1:T$

Train weak learning algorithm using data weighted by d_{ti} . This produces weak classifier h_t .

Choose coefficient α_t .

$y_t h_t(x_i)$ is 1 if correct, -1 if incorrect

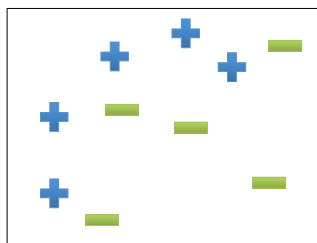
Update weights:

$$d_{t+1,i} = \frac{d_{t,i} \exp(-\alpha_t y_t h_t(x_i))}{Z_t} \quad Z_t \text{ is a normalization factor.}$$

End

Output the final classifier: $H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x_i)\right)$

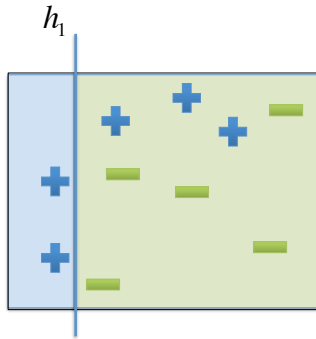
Boosting Example



All points start with equal weights.

(Credit: Example adapted from Freund and Schapire)

Boosting Example

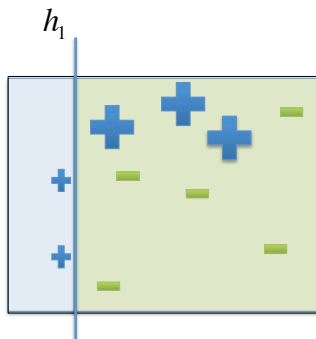


Run the weak learning algorithm to get a weak classifier.

Choose coefficient $\alpha_1 = .41$

(Credit: Example adapted from Freund and Schapire)

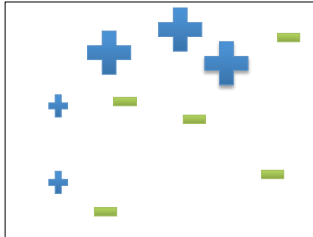
Boosting Example



Increase the weights on the misclassified points, decrease the weights on the correctly classified points.

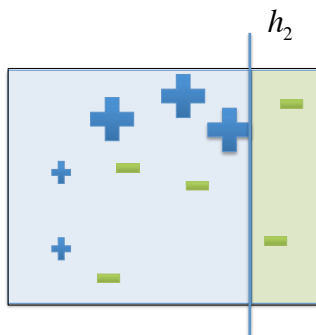
(Credit: Example adapted from Freund and Schapire)

Boosting Example



(Credit: Example adapted from Freund and Schapire)

Boosting Example

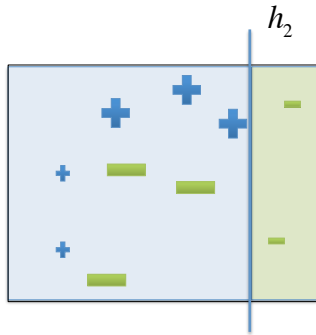


Run the weak learning algorithm to get a weak classifier for the weighted data.

Choose coefficient $\alpha_2 = .66$

(Credit: Example adapted from Freund and Schapire)

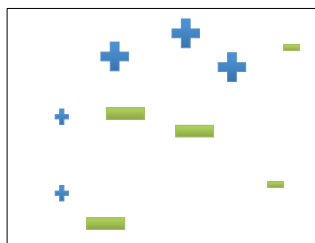
Boosting Example



Increase the weights on the misclassified points, decrease the weights on the correctly classified points.

(Credit: Example adapted from Freund and Schapire)

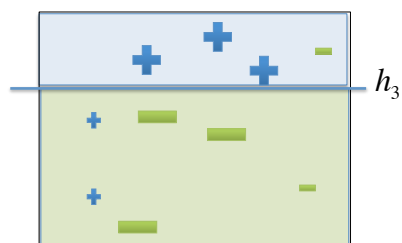
Boosting Example



Increase the weights on the misclassified points, decrease the weights on the correctly classified points.

(Credit: Example adapted from Freund and Schapire)

Boosting Example

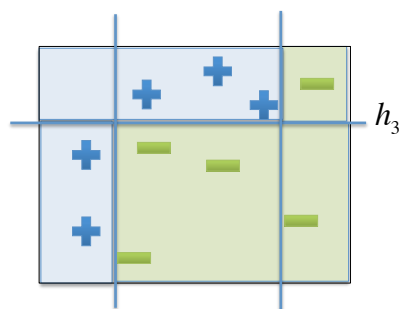


Increase the weights on the misclassified points, decrease the weights on the correctly classified points.

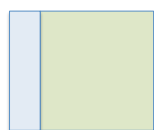
Choose coefficient $\alpha_3 = .93$

(Credit: Example adapted from Freund and Schapire)

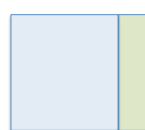
Boosting Example



$H = \text{sign}(.42$



$+ .66$



$+ .93$



$)$

(Credit: Example adapted from Freund and Schapire)