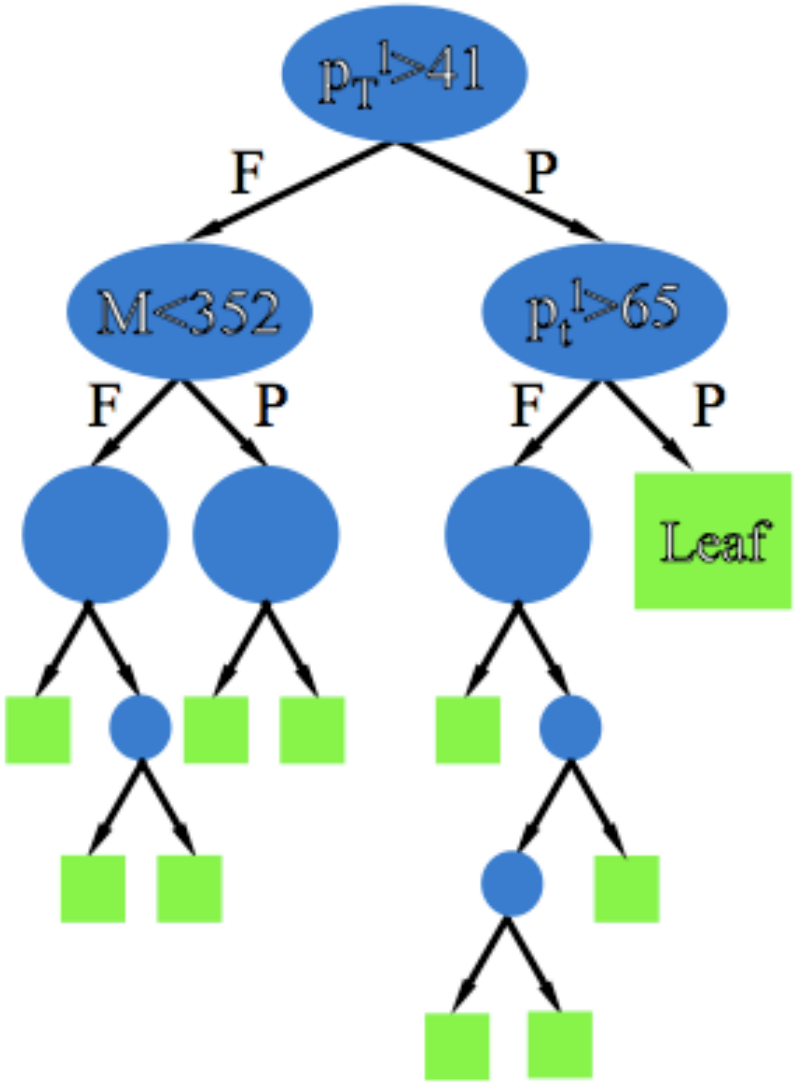
Advanced Analysis Methods

Tulika Bose

April 29th, 2016

Trees and leafs



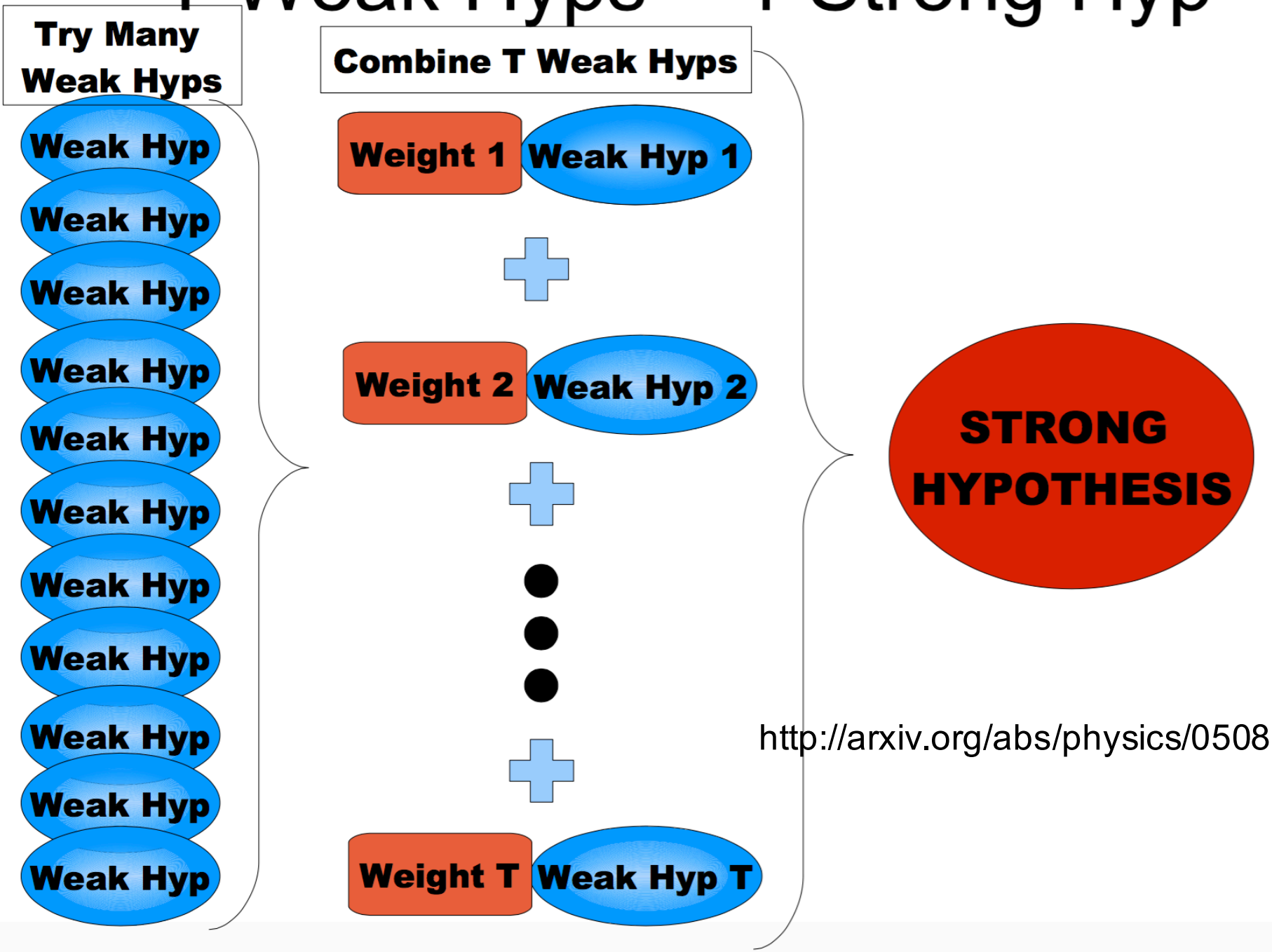
- Create a tree of cuts
- Divide sample into "pass" and "fail" sets
- Each node corresponds to a cut (branch)
- A leaf corresponds to an end-point
- For each leaf, calculate purity (from MC):
  purity = $N_S/(N_S+N_B)$

Repeat recursively on each node

Stop (terminate at leaf) when improvement stops or when too few events left

Boosting

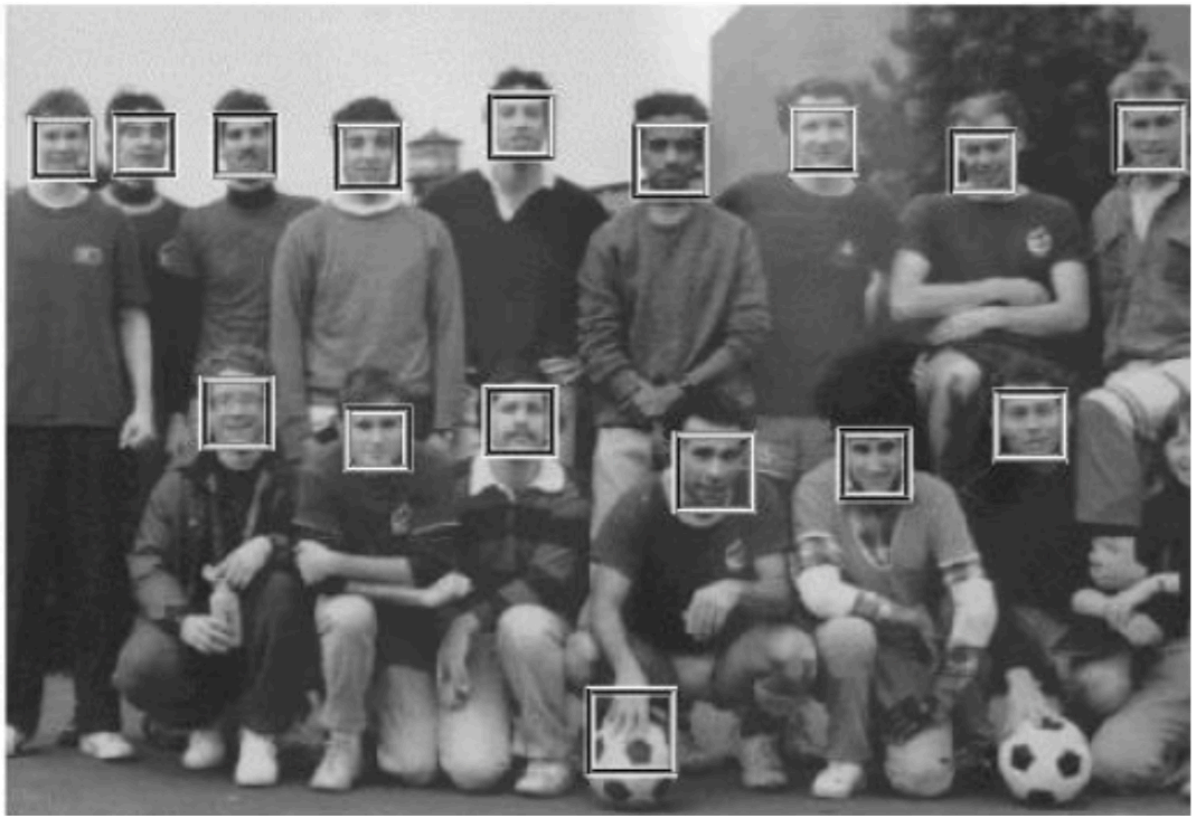
T Weak Hyps = 1 Strong Hyp



http://arxiv.org/abs/physics/0508045

Example: Face Detection

- We are given a dataset of images
- We need to determine if there are faces in the images



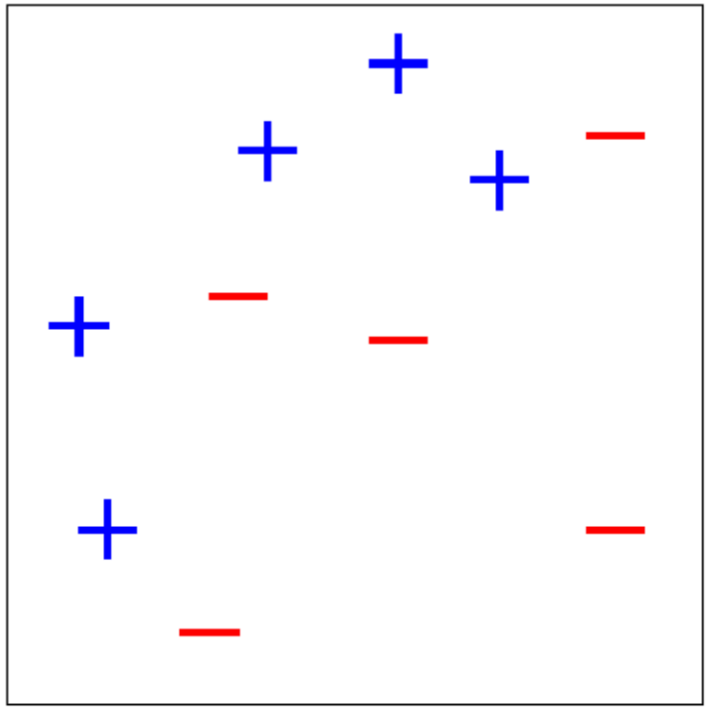
Example: Face Detection

- Go through each possible rectangle
- Some weak hypotheses might be:
  - Is there a round object in the rectangle?
  - Does the rectangle have darker spots where the eyes should be?
  - Etc.
- Classifier = 2.1 * (Is Round) + 1.2 * (Has Eyes)

Viola and Jones, 2011

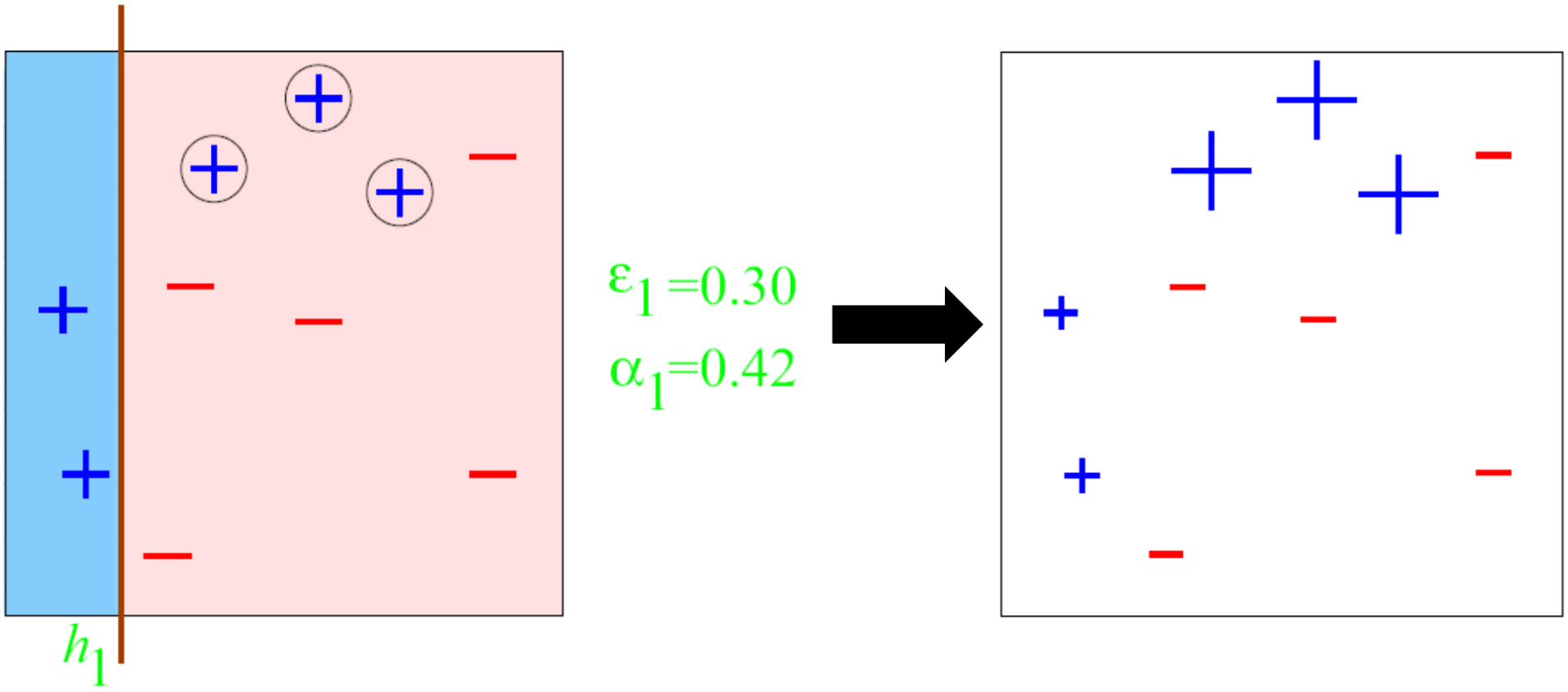
Boosting: Toy Example

- Positive examples
- Negative examples
- 2-Dimensional plane
- Weak hyps: linear separators
- 3 iterations



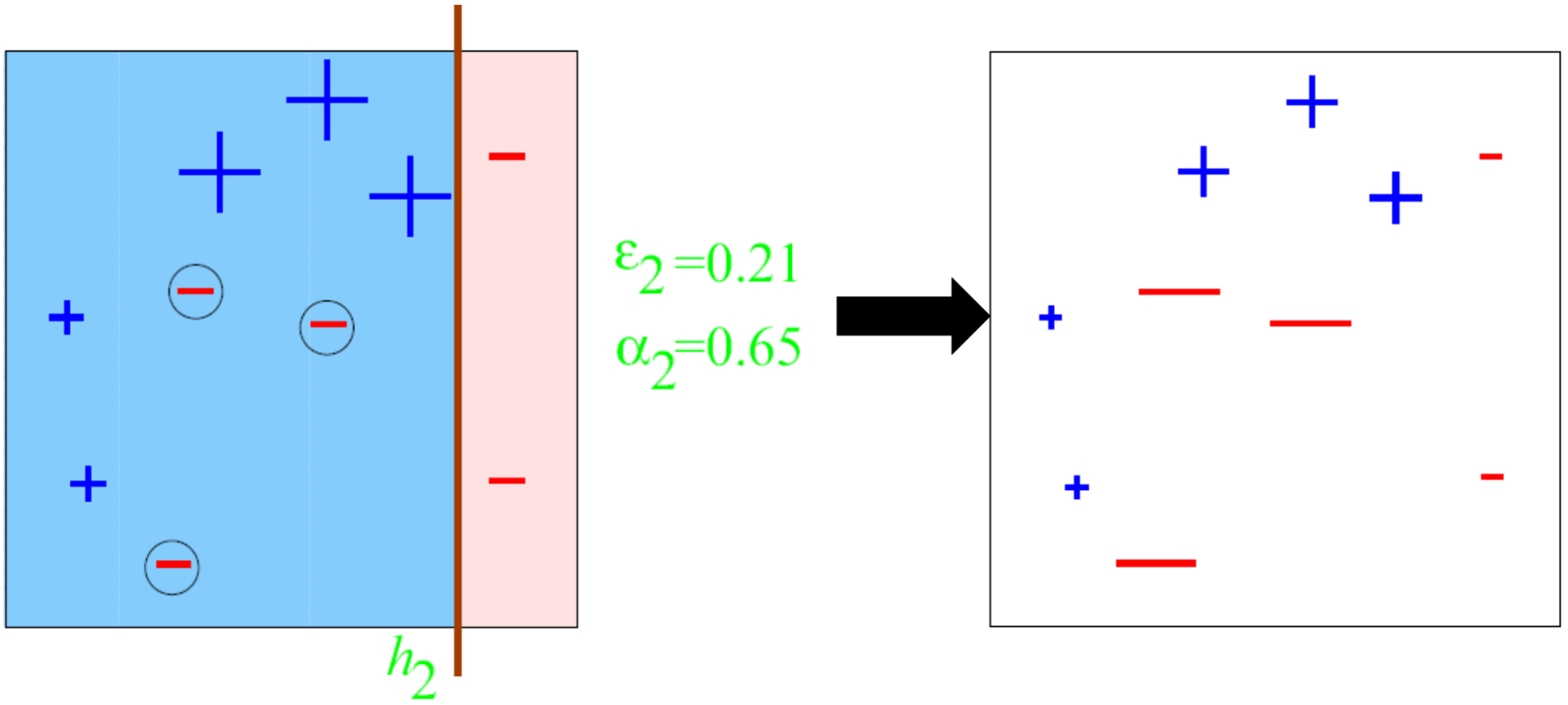
Freund, 1996

Toy Example: Iteration 1



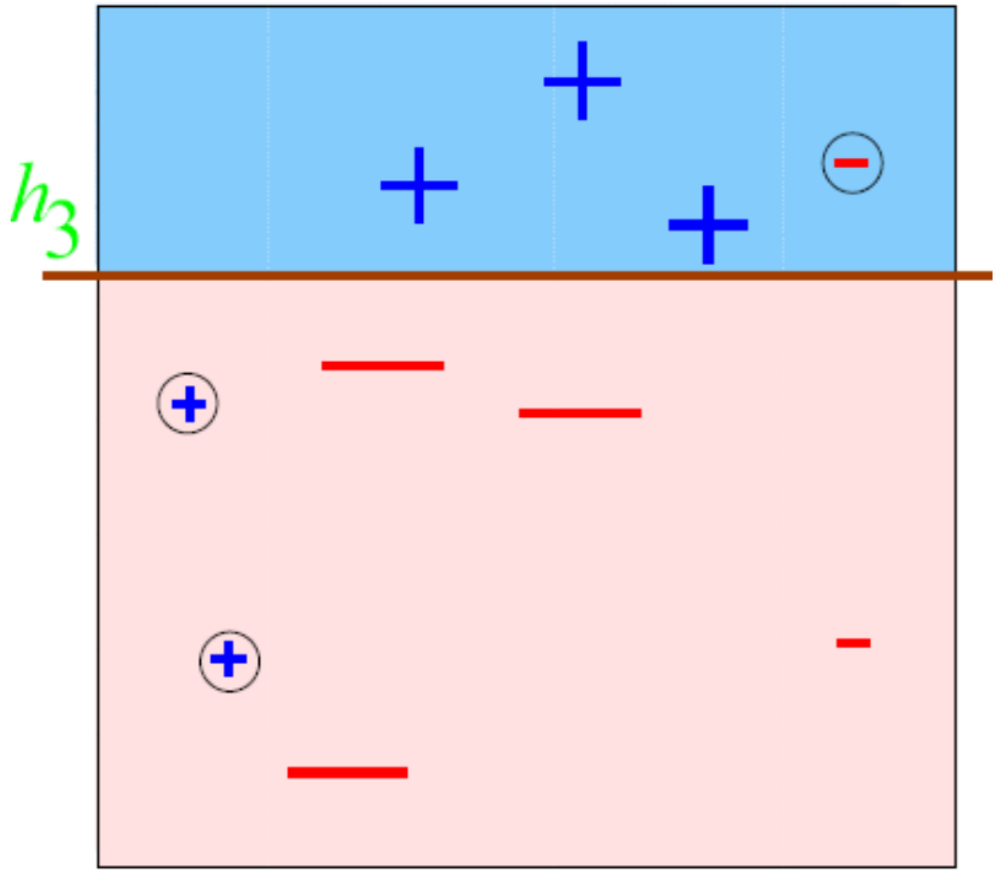
Misclassified examples are circled, given more weight

Toy Example: Iteration 2



Misclassified examples are circled, given more weight

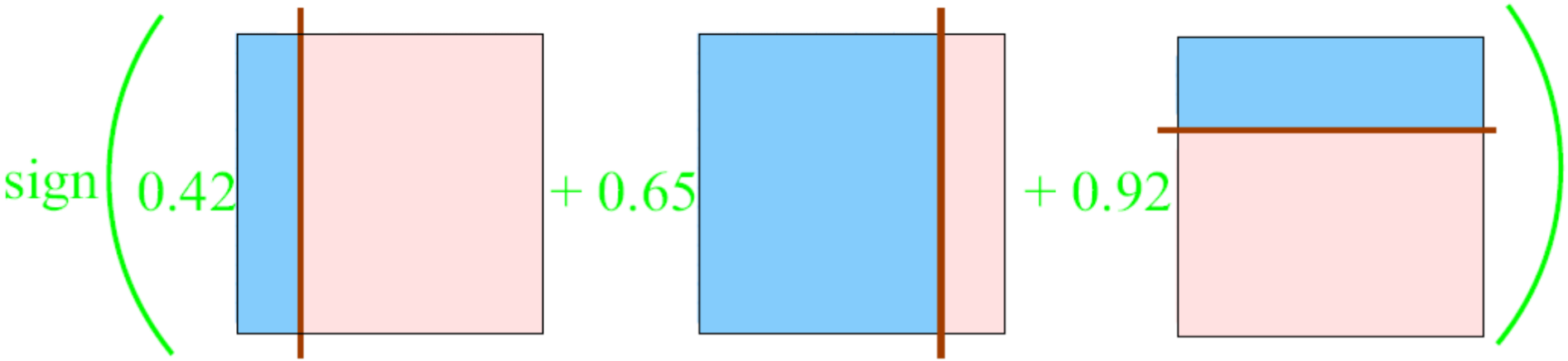
Toy Example: Iteration 3

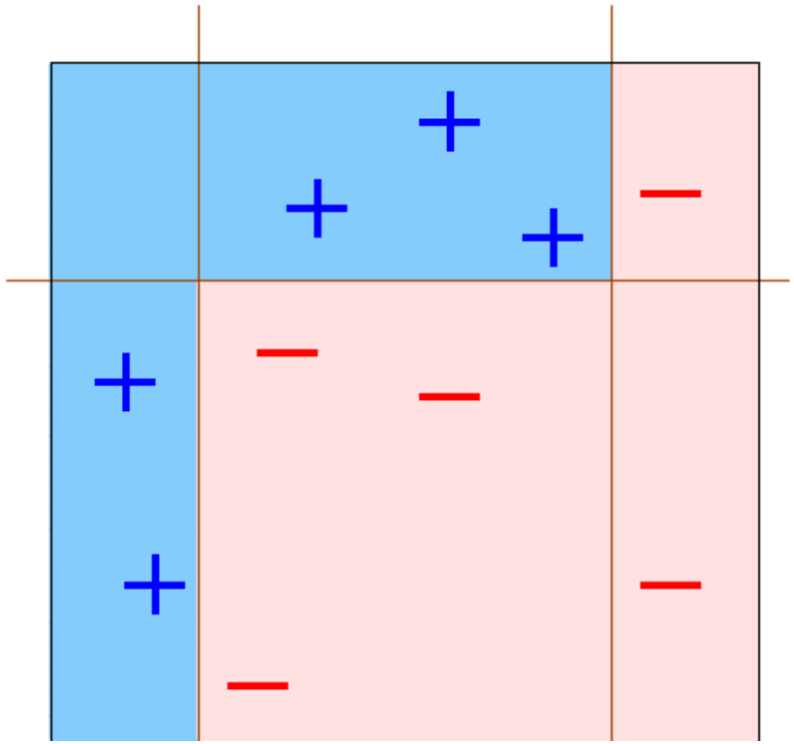


$\varepsilon_3 = 0.14$

$\alpha_3 = 0.92$

Finished boosting



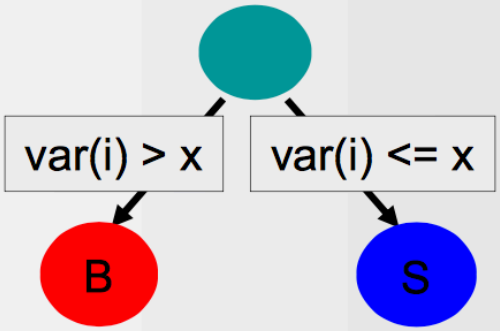


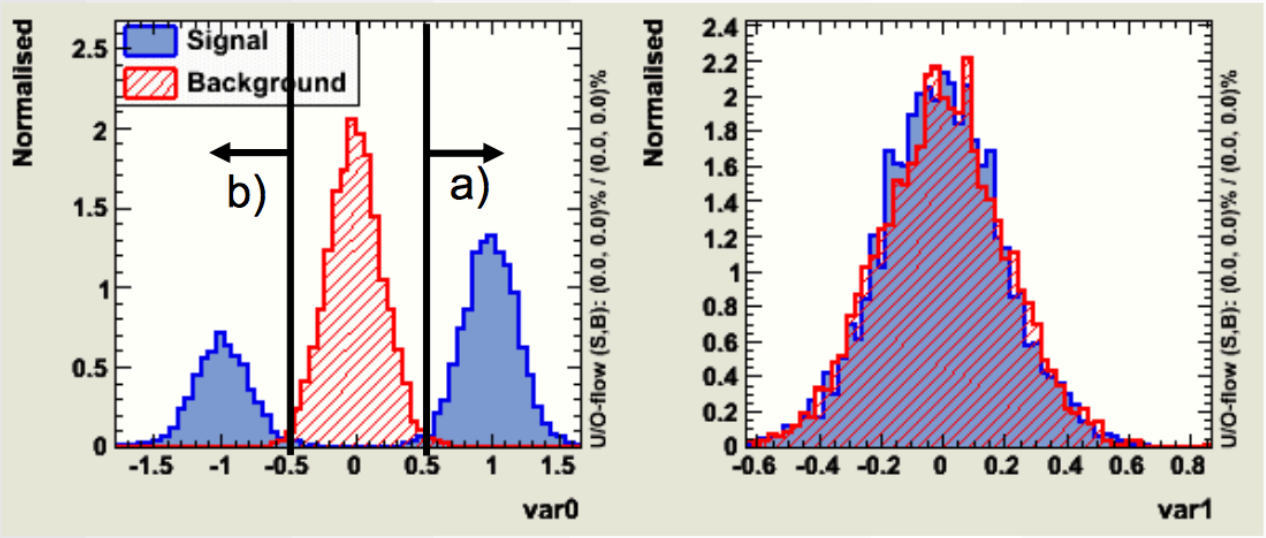
Taken from Freund 1996

AdaBoost: A simple demonstration

The example: (somewhat artificial…but nice for demonstration) :

- Data file with three “bumps”
- Weak classifier (i.e. one single simple “cut” ↔ decision tree stumps)





Two reasonable cuts: a) Var0 > 0.5 → ε_{signal}=66% ε_{bkg} ≈ 0% misclassified events in total 16.5%

or

b) Var0 < -0.5 → ε_{signal}=33% ε_{bkg} ≈ 0% misclassified events in total 33%

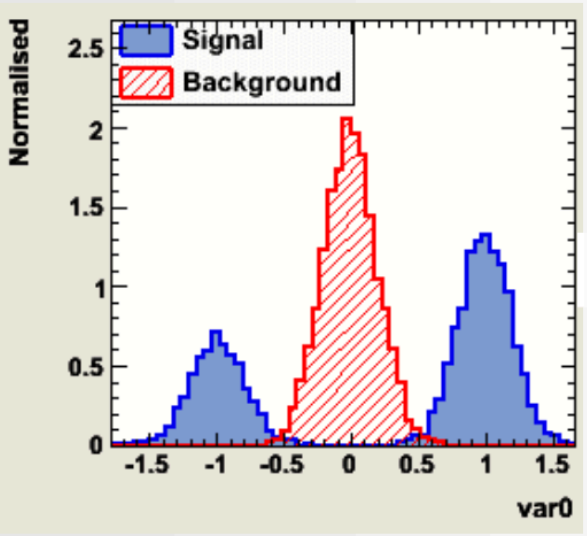
the training of a single decision tree stump will find “cut a)”

AdaBoost: A simple demonstration

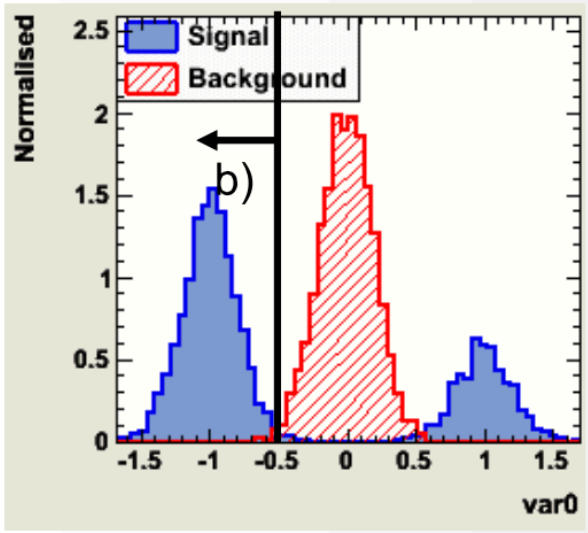
The first “tree”, choosing cut a) will give an error fraction: err = 0.165

before building the next “tree”: weight wrong classified training events by (1-err)/err ≈ 5

the next “tree” sees essentially the following data sample:



re-weight

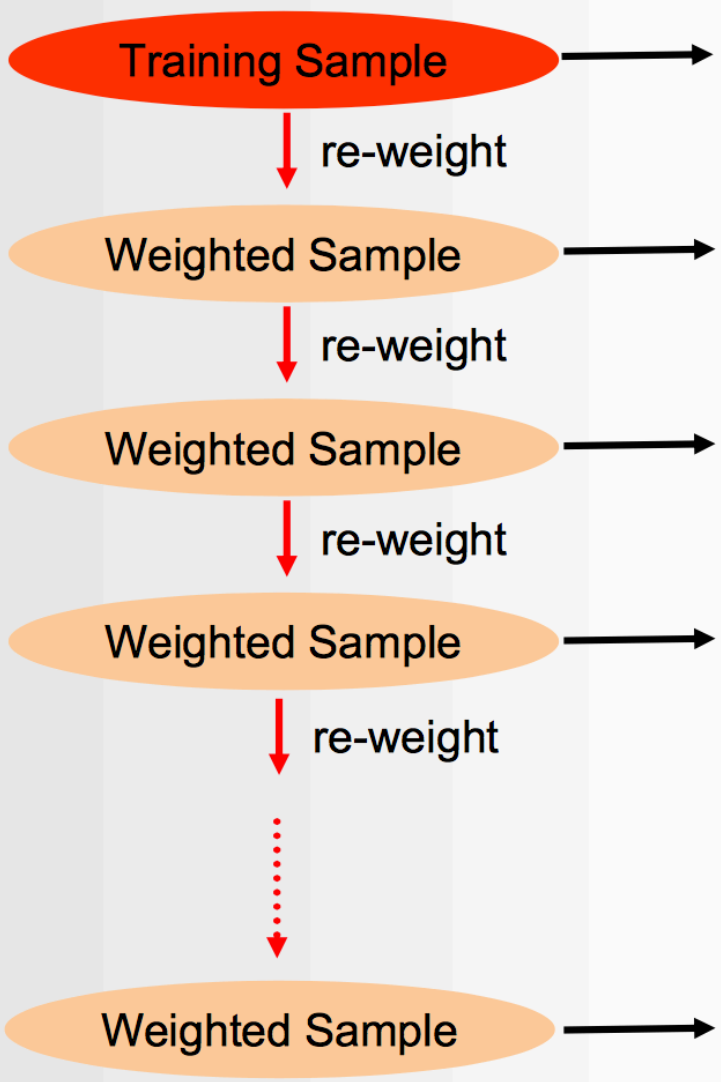


.. and hence will chose: “cut b)”: Var0 < -0.5

The combined classifier: Tree1 + Tree2

the (weighted) average of the response to a test event from both trees is able to separate signal from background as good as one would expect from the most powerful classifier

Adaptive Boosting (AdaBoost)



- AdaBoost re-weights events misclassified by previous classifier by:

$$\frac{1-f_{err}}{f_{err}} \quad \text{with :}$$

$$f_{err} = \frac{\text{misclassified events}}{\text{all events}}$$

- AdaBoost weights the classifiers also using the error rate of the individual classifier according to:

$$y(x) = \sum_{i}^{N_{Classifier}} \log\left(\frac{1-f_{err}^{(i)}}{f_{err}^{(i)}}\right) C^{(i)}(x)$$