CSCM77
Haar & Adaboost for Face Detection

Prof. Xianghua Xie
x.xie@swansea.ac.uk
http://csvision.swan.ac.uk

Face Detection

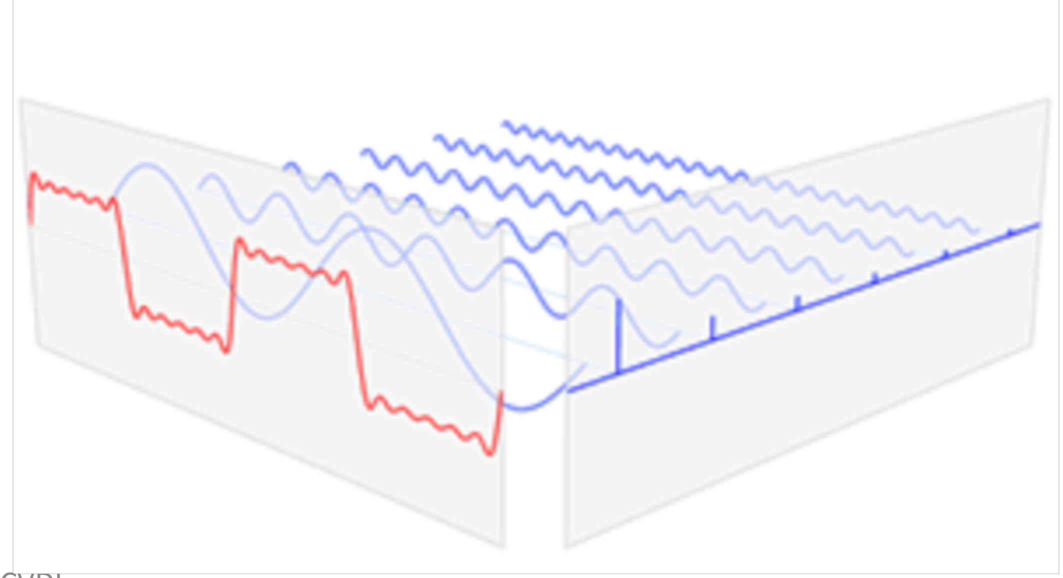
- To perform face detection, we introduce two (not entirely) new concepts
- Wavelets
- AdaBoost

Recap: Edge Detection

- Use derivative operators to capture and measure intensity discontinuity, e.g.
  - 1^{st} order derivatives: Prewitt & Sobel;
    - Prewitt: $M_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$ $M_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$
    - Sobel: $M_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$ $M_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$
  - 2^{nd} order derivative: LoG
- The key is to evaluate spatial discontinuity
- Those filters generally do not take into account scale changes
  - That is the filters are fixed in size

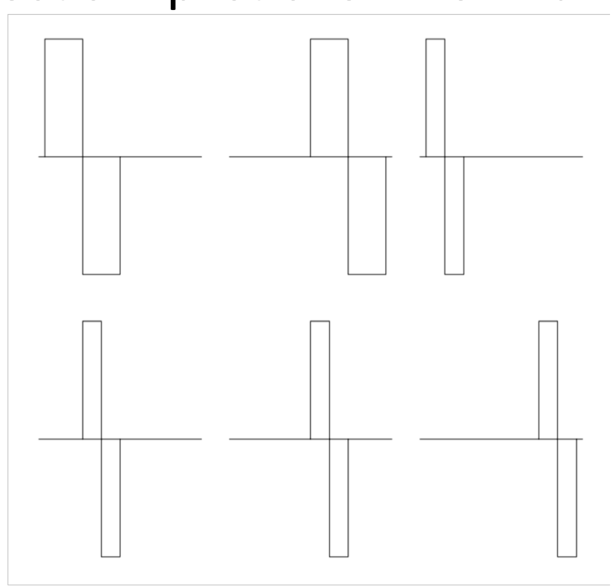
Analysis in frequency domain

- Fourier analysis
  - From Fourier theory that a signal can be expressed as the sum of a, possibly infinite, series of sines and cosines.
    - This sum is also referred to as a Fourier expansion.
  - Fourier basis are localised in frequency but not in space (or time, in the case of dealing with time-varying signals)
    - Small changes in Fourier domain creates changes everywhere in spatial domain.



Haar wavelet

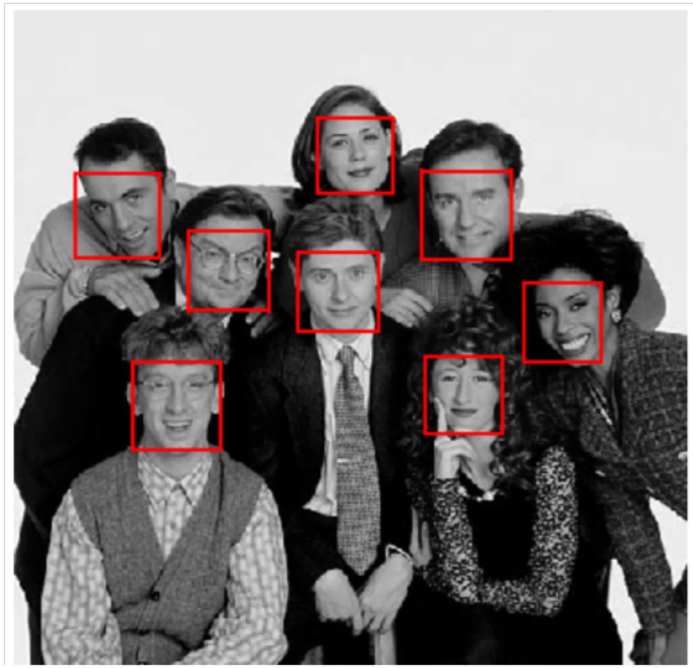
- Wavelets
  - "wave" like oscillating functions; integral equals to zero
  - A self-similar family of functions
  - Wavelets are local in both frequency domain and in spatial domain
    - Dilation allows localisation in frequency domain
    - Translation allows localisation in spatial domain
  - Haar is the most simplistic form of wavelets

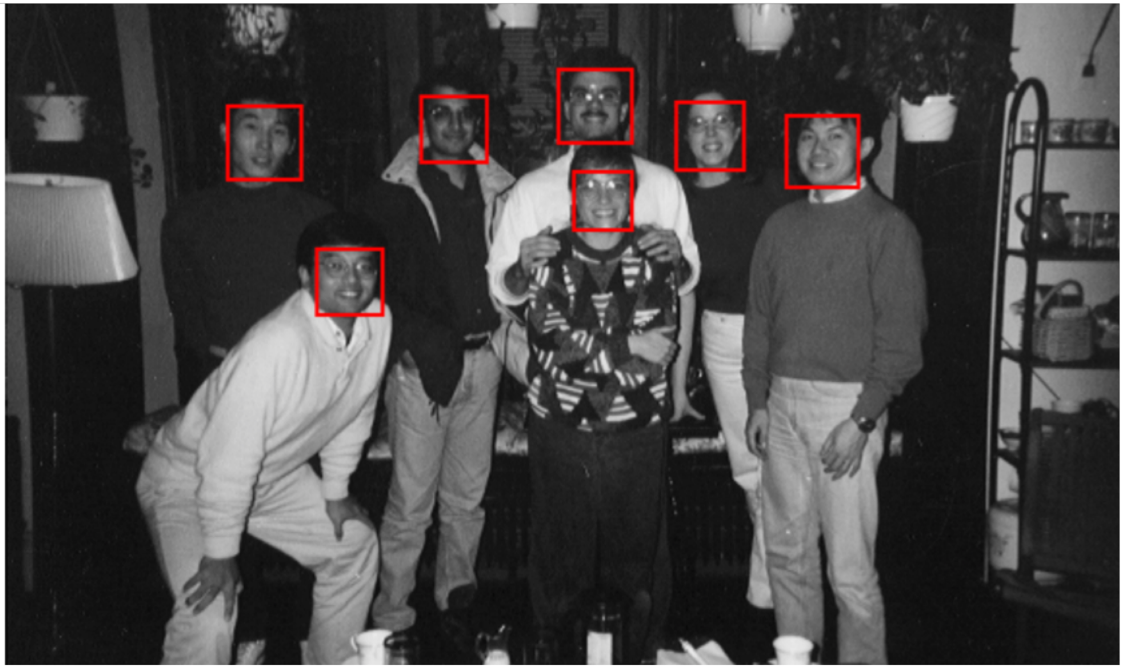


Dilation and translation of Haar wavelet on [0,1]

Face Detection

- Similar to human detection, sliding windows detectors are used to evaluate tens of thousands of location and scale combinations
  - Thus, evaluation needs to be computationally efficient
    - This requires feature extraction to be efficient
    - And the classification needs to be efficient as well





Face Detection

- Viola-Jones Face Detector
  - Training is relatively slow, but detection is very fast

- Key ideas
  - Haar-like image features
  - Integral images for fast feature extraction (not discussed in this lecture)
  - Boosting for feature selection
  - Attentional cascade for fast rejection of non-face hypotheses

P. Viola and M. Jones. *Robust real-time face detection.* IJCV 57(2), 2004.

Image features

- Haar-like filters
  - Rectangular filters consisting of +1, -1 coefficients
  - Appropriate combinations of added and subtracted rectangles approximate various (degree) derivative filters, at different locations and scales

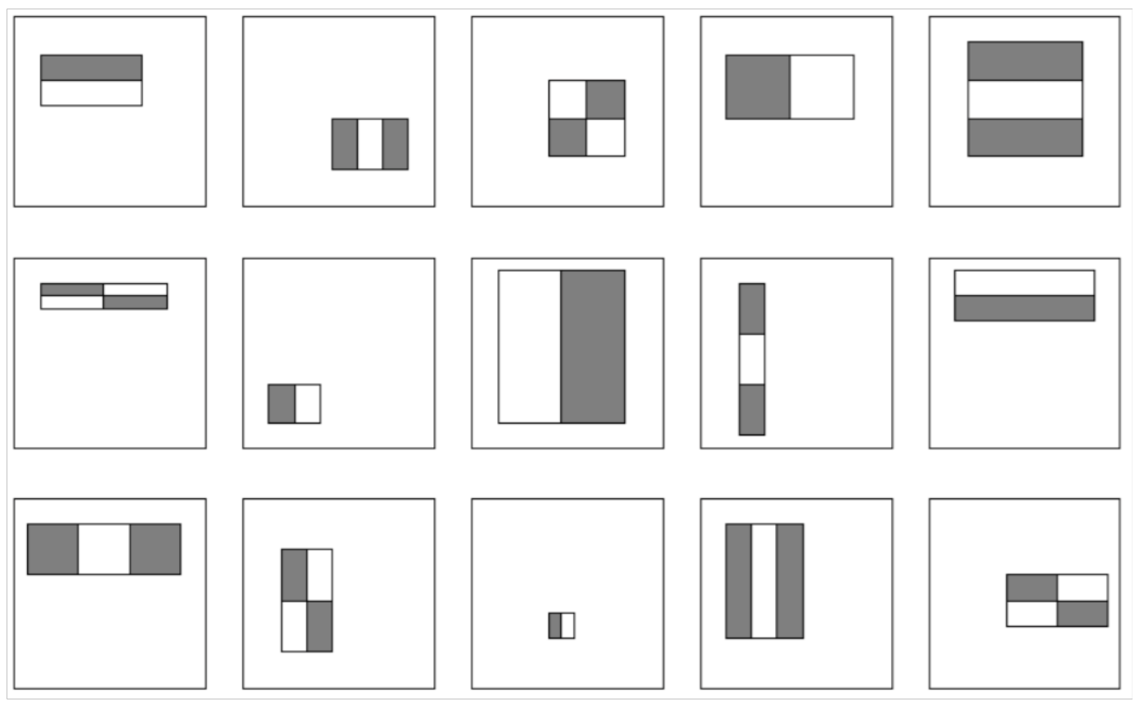
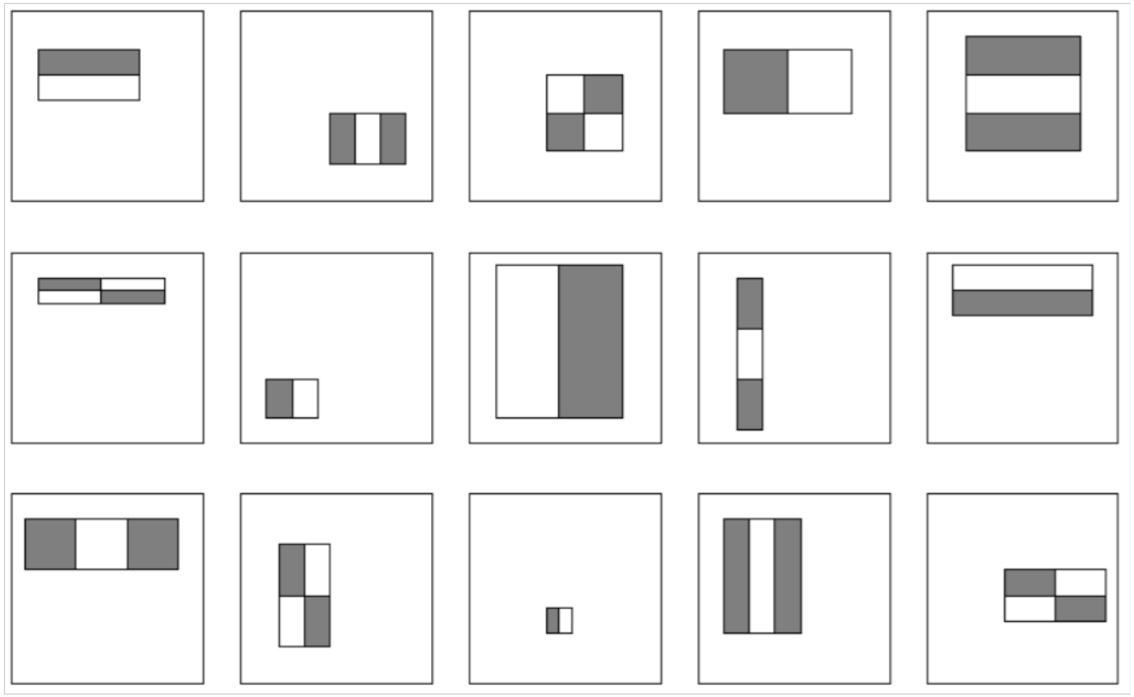


Image features

- Efficiency considerations
  - Large number of filters: for a typical 24x24 detection window, the number of Haar features considered is ~160,000
  - At test time, it is impractical to evaluate all features
  - We thus need to build a good/strong classifier using only a small fraction of the features

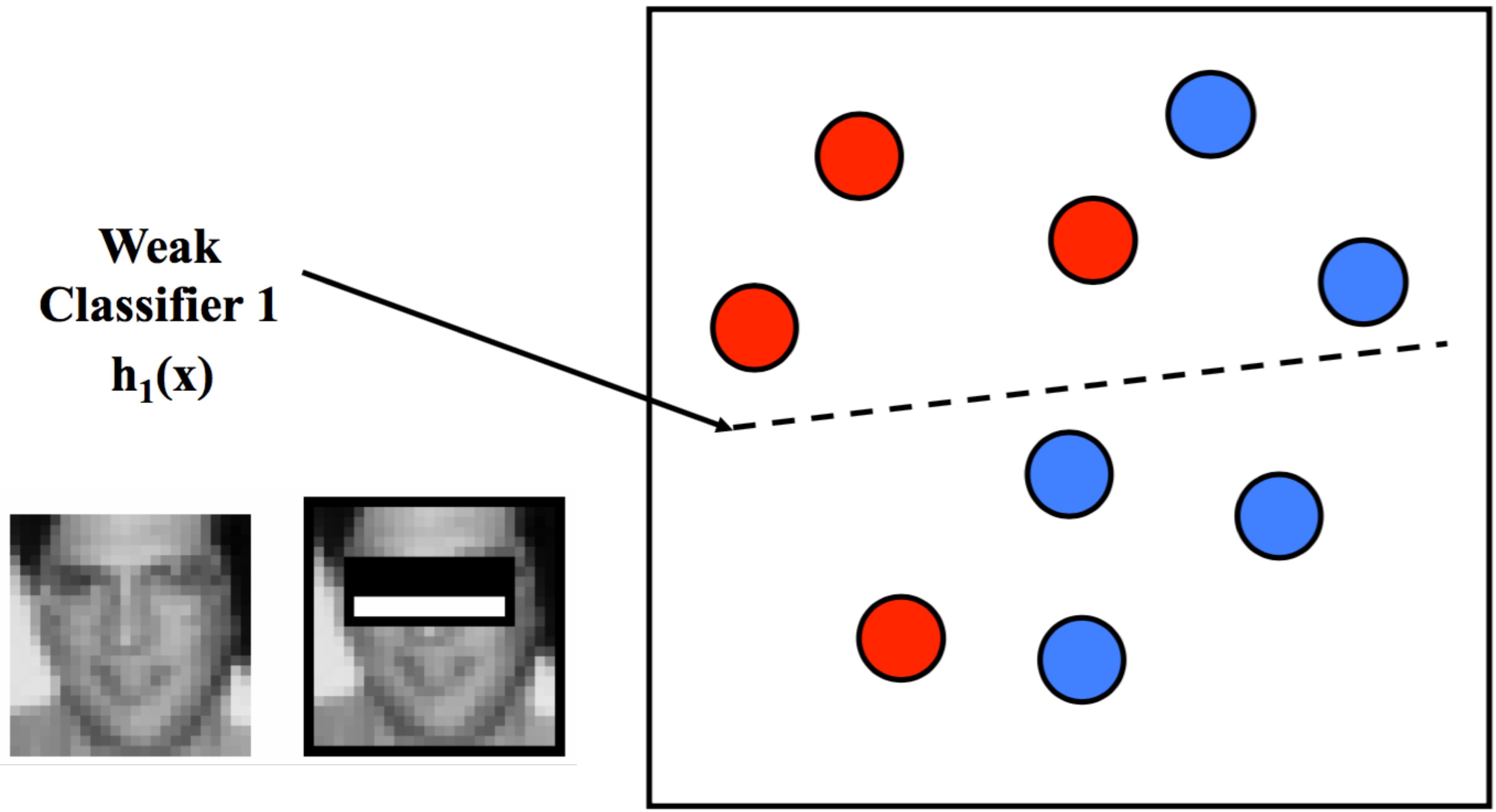


Boosting

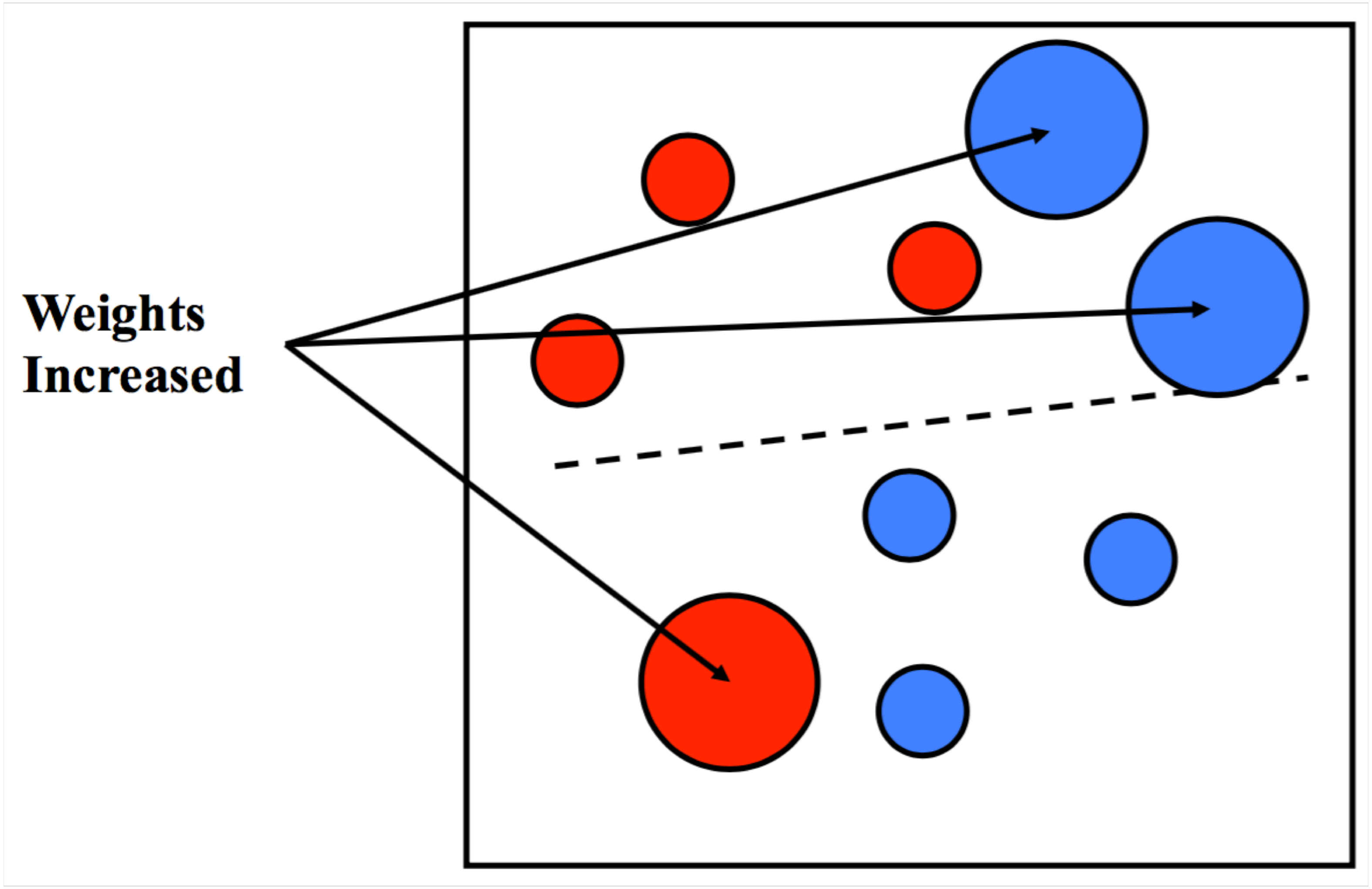
- Boosting is a classification scheme that works by combining weak learners into a more accurate ensemble classifier
  - A weak learner need only do better than chance

- Training consists of multiple boosting rounds
- During each boosting round, we select a weak learner that does well on examples that were hard for the previous weak learners
- "Hardness" is captured by weights attached to training examples and used in select the next weak learner
  - AdaBoost: adaptive boosting is a good example of boosting

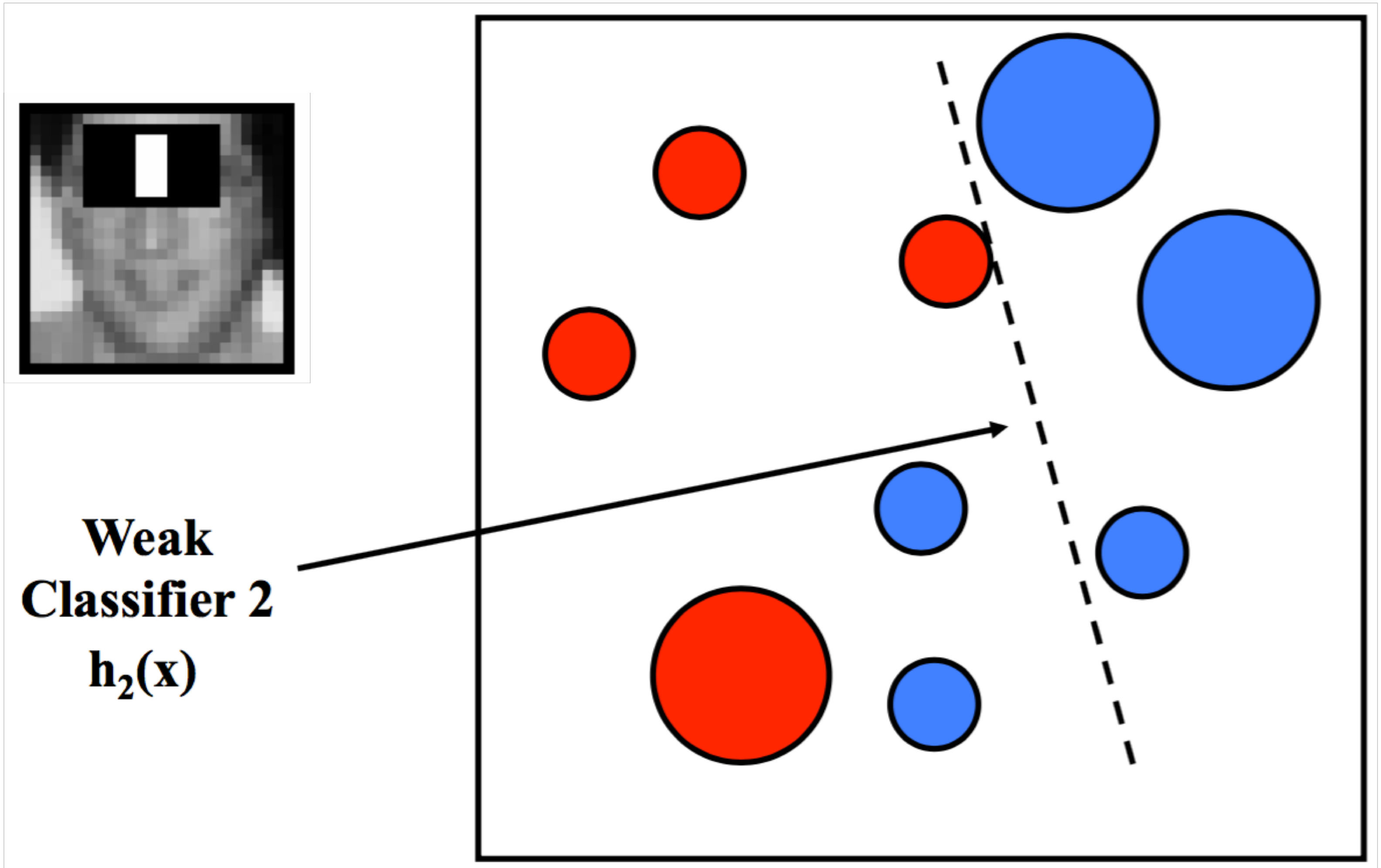
Boosting



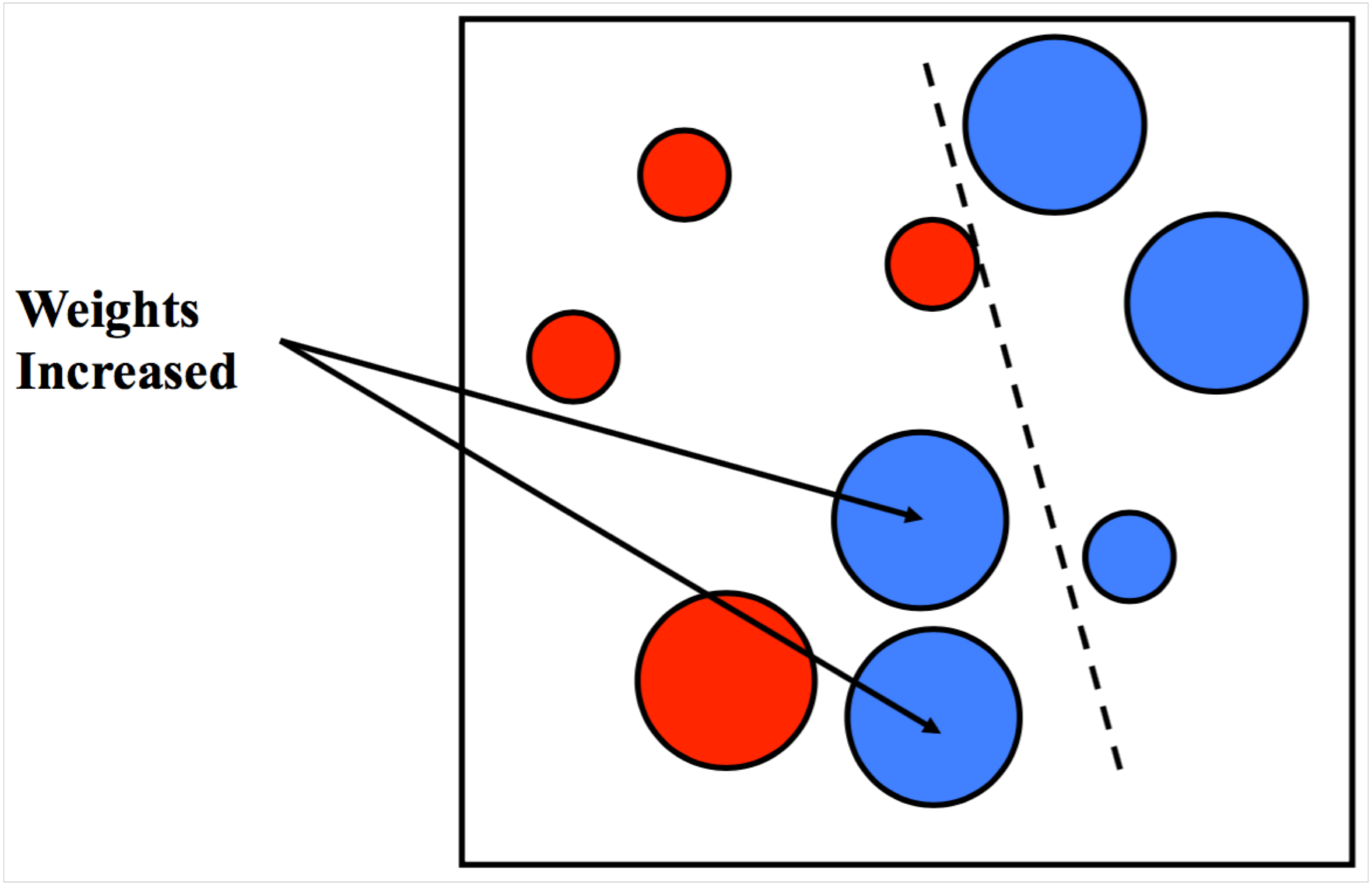
Boosting



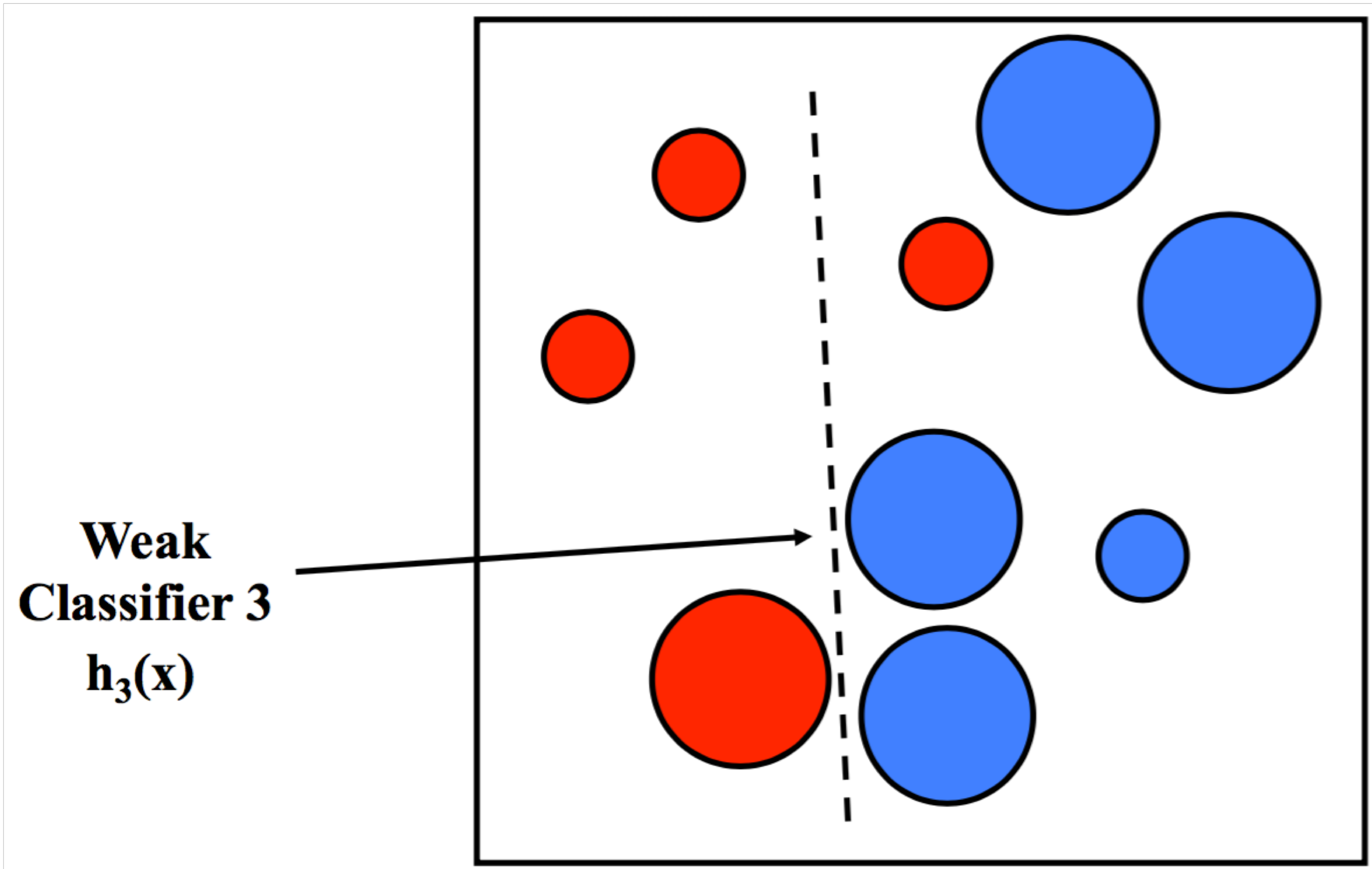
Boosting



Boosting



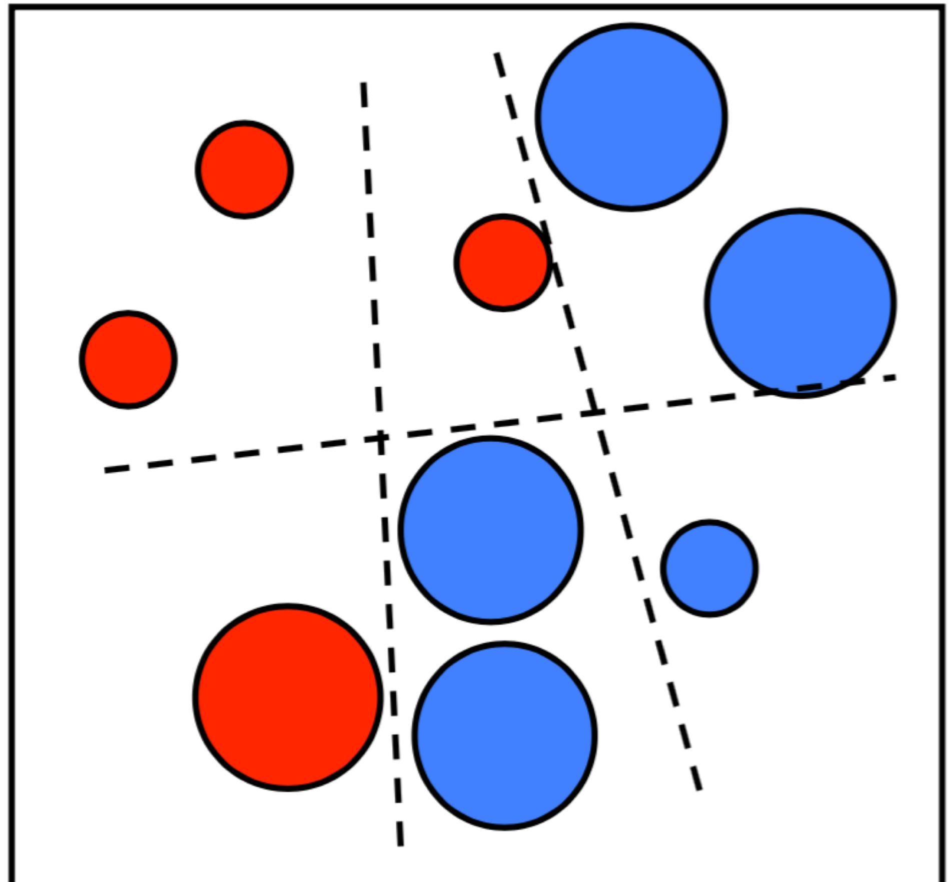
Boosting



Boosting

Final classifier is a weighted combination of the weak classifiers

$$h(\mathbf{x}) = \operatorname{sign}\left(\sum_{j=1}^{M} \alpha_j h_j(\mathbf{x})\right)$$

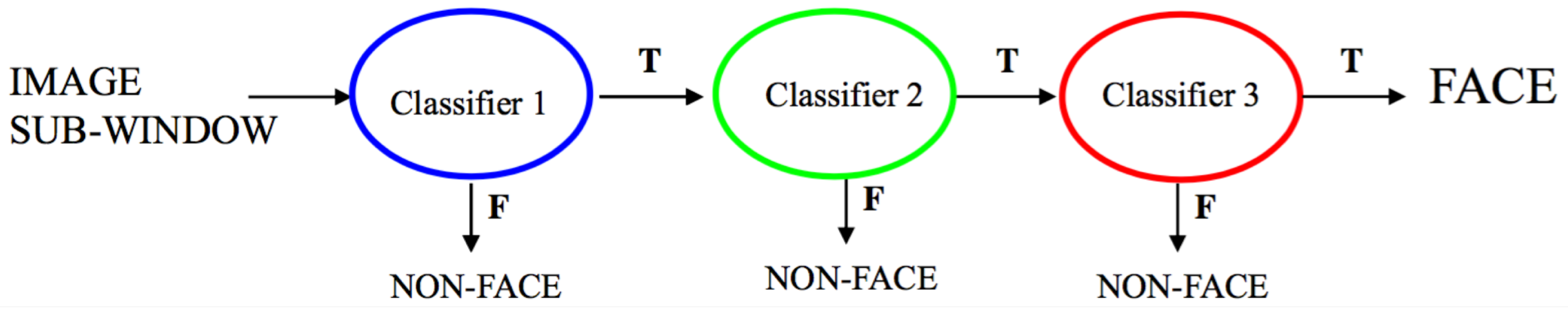


Boosting

- AdaBoost attempts to boost the accuracy of an ensemble of weak classifiers.
  - Each weak classifier is trained stage-wise to minimise the empirical error for a given distribution reweighted according to the classification errors of the previously trained classifiers.
  - It is shown that AdaBoost is a sequential forward search procedure using the greedy selection strategy to minimise a certain margin on the training set.
- A crucial heuristic assumption used in such a sequential forward search procedure is the monotonicity (i.e. that addition of a new weak classifier to the current set does not decrease the value of the performance criterion).
- The premise offered by the sequential procedure in AdaBoost breaks down when this assumption is violated.

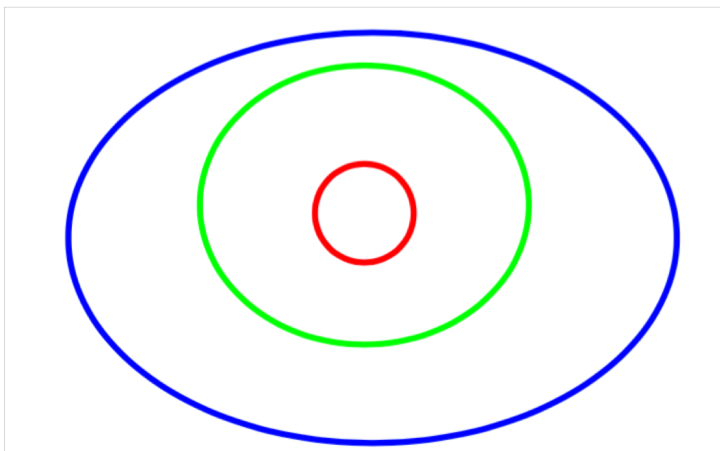
Cascading

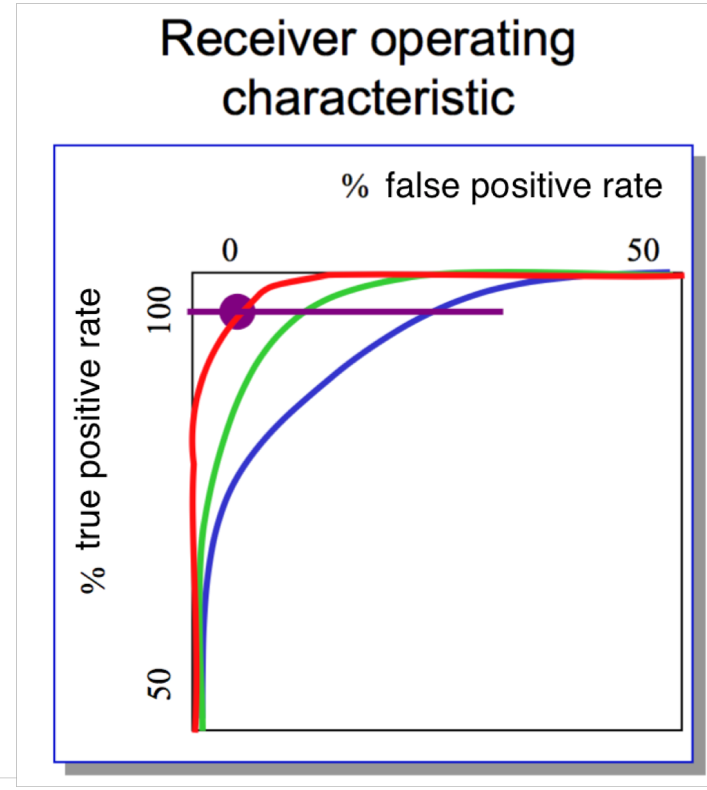
- We start with simple classifiers which reject many of the negative sub-windows while detecting almost all positive sub-windows
- Positive results from the first classifier triggers the evaluation of a second (more complex) classifier, and so on
- A negative outcome at any point leads to the immediate rejection of the sub-window

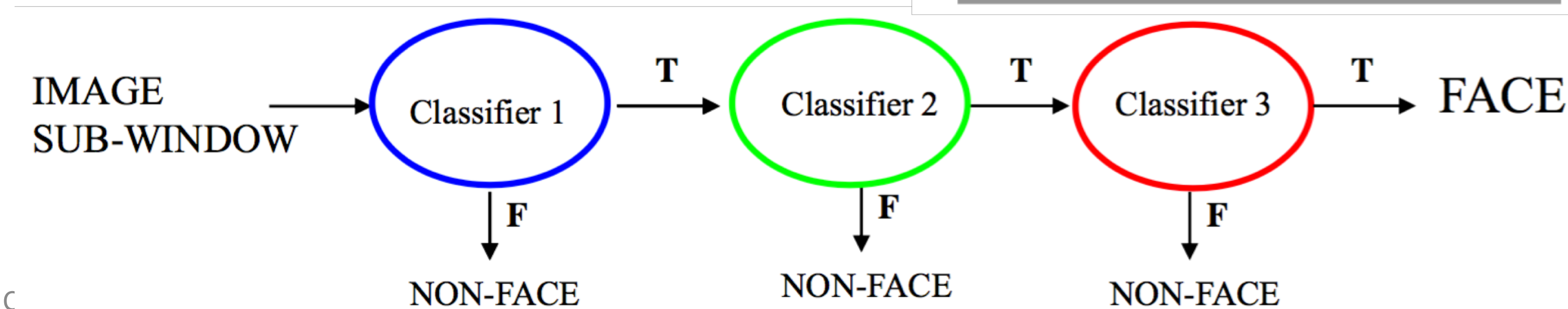


Cascading

- Chain classifiers that are progressively more complex and have lower false positive rates:







Example implementation

- Training Data
  - 5000 faces
  - All frontal, rescaled to 24x24 pixels
  - 300 million non-faces
  - 9500 non-face images
  - Faces are normalised: scale & translation
- Total of 6061 features
- Average of 10 features evaluated per window on test set
  - Real time

Example implementation