Boosting and Viola Jones
Application: Face Detection (Viola-Jones’00)

Given a rectangular window of pixels, is there a face in it?

Properties:
* Easy to come up with simple rules-of-thumb classifiers,
* Hard to come up with a single high accuracy rule
Viola-Jones Weak Learners

A weak learner $h_{f,t,s}$ is described by:
* feature $f$
* threshold $t$
* sign $s$ (+1 or -1)

For an example $x$,

$$h_{f,t,s}(x) = 1, \text{ if } sf(x) \geq t$$
$$= -1, \text{ otherwise}$$
Viola-Jones: 3 Types of Features

1. Feature value = sum of pixel colors in black rectangle - sum of pixel colors in white rectangle

2. Feature value = sum of pixel colors in black rectangles - sum of pixel colors in white rectangle

3. Feature value = sum of pixel colors in black rectangles - sum of pixel colors in white rectangles
Viola-Jones Weak Learners

A weak learner $h_{f,t,s}$ is described by:
* feature $f$ (3 types of rectangular features)
* threshold $t$
* sign $s$ (+1 or -1)

For an example $x$,
\[
h_{f,t,s}(x) = \begin{cases} 
1, & \text{if } sf(x) \geq t \\
-1, & \text{otherwise}
\end{cases}
\]
Viola-Jones: Computing the Features

1

2

3
Precompute and store the values \( s(x, y) \) for each \((x, y)\):

\[
\begin{array}{c}
(x, y) \\
\end{array}
\]

\( s(x, y) = \text{sum of pixel colors in the black rectangle} \)

Now each feature can be computed from adding/subtracting a constant number of \( s(x, y) \)'s
Viola-Jones: Procedure

Given training set $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, $y$ in $\{-1, 1\}$

For $t = 1, \ldots, T$

- Construct distribution $D_t$ on the examples
- Find weak learner $h_t$ which has small error $\text{err}_{D_t}(h_t)$ wrt $D_t$

Output final classifier

Weak learning procedure: Find the feature $f$, sign $s$, and threshold $t$ for which the error of $h_{f,t,s}$ on $D_t$ is minimum
Viola-Jones: Procedure

Given training set $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, $y$ in $\{-1, 1\}$

For $t = 1, \ldots, T$

Construct distribution $D_t$ on the examples

Find weak learner $h_t$ which has small error $\text{err}_{D_t}(h_t)$ wrt $D_t$

Output final classifier

Initially, $D_1(i) = 1/n$, for all $i$ (uniform)

Given $D_t$ and $h_t$:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \exp(-\alpha_t y_i h_t(x_i))$$

where:

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \text{err}_{D_t}(h_t)}{\text{err}_{D_t}(h_t)} \right)$$

Final classifier: $\text{sign} \left( \sum_{t=1}^{T} \alpha_t h_t(x) \right)$
Viola and Jones: Some Results

Below $X$ is uniform over $[0, 1]$, and $Y = 1$ if $X > 0.5$, 0 otherwise.
Cascades for Fast Classification

Choose thresholds for low false negative rates
Fast classifiers earlier in cascade, slower classifiers later
Most examples don’t get to the later stages, so system is fast on average