Boosting

• Defines a classifier using an additive model:

\[ F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots \]
Boosting

- Defines a classifier using an additive model:

\[ F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots \]

- We need to define a family of weak classifiers

\[ f_k(x) \text{ from a family of weak classifiers} \]
Each data point has a class label: \( y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases} \) and a weight: \( w_t = 1 \)

• It is a sequential procedure:
Toy example

Weak learners from the family of lines

Each data point has a class label:

\[ y_t = \begin{cases} 
+1 & (\bigcirc) \\
-1 & (\bigotimes) 
\end{cases} \]

and a weight:

\[ w_t = 1 \]

\[ h \Rightarrow p(\text{error}) = 0.5 \text{ it is at chance} \]
This one seems to be the best

Each data point has a class label:

\[ y_t = \begin{cases} 
  +1 & (\bigcirc) \\
  -1 & (\bigotimes) 
\end{cases} \]

and a weight:

\[ w_t = 1 \]

This is a ‘\textbf{weak classifier}’: It performs slightly better than chance.
We set a new problem for which the previous weak classifier performs at chance again.

Each data point has a class label:

\[ y_t = \begin{cases} +1 & (\bullet) \\ -1 & (\circ) \end{cases} \]

We update the weights:

\[ w_t \leftarrow w_t \exp\{-y_t H_t\} \]
We set a new problem for which the previous weak classifier performs at chance again.

Each data point has a class label:

\[ y_t = \begin{cases} +1 & (\bullet) \\ -1 & (\circ) \end{cases} \]

We update the weights:

\[ w_t \leftarrow w_t \exp\{-y_t H_t\} \]
We set a new problem for which the previous weak classifier performs at chance again.

Each data point has a class label:
\[ y_t = \begin{cases} 
+1 & (\circ) \\
-1 & (\bigcirc)
\end{cases} \]

We update the weights:
\[ w_t \leftarrow w_t \exp\{-y_t H_t\} \]
We set a new problem for which the previous weak classifier performs at chance again.

Each data point has a class label:

\[ y_t = \begin{cases} 
+1 \ (\circ) \\
-1 \ (\lozenge) 
\end{cases} \]

We update the weights:

\[ w_t \leftarrow w_t \exp\{-y_t H_t\} \]
The strong (non-linear) classifier is built as the combination of all the weak (linear) classifiers.
Boosting

• Different cost functions and minimization algorithms result in various flavors of Boosting.
• In this demo, I will use gentleBoosting: it is simple to implement and numerically stable.
Overview of section

• Object detection with classifiers

• Boosting
  – *Gentle boosting*
  – Weak detectors
  – Object model
  – Object detection
Boosting fits the additive model

\[ F(x) = f_1(x) + f_2(x) + f_3(x) + \ldots \]

by minimizing the exponential loss

\[ J(F) = \sum_{t=1}^{N} e^{-y_t F(x_t)} \]

Training samples

The exponential loss is a differentiable upper bound to the misclassification error.
Exponential loss

Misclassification error
Squared error
Exponential loss

Squared error

\[ J = \sum_{t=1}^{N} \left[ y_t - F(x_t) \right]^2 \]

Exponential loss

\[ J = \sum_{t=1}^{N} e^{-y_tF(x_t)} \]
Boosting

Sequential procedure. At each step we add

$$F(x) \leftarrow F(x) + f_m(x)$$

to minimize the residual loss

$$(\phi_m) = \arg \min_{\phi} \sum_{t=1}^{N} J(y_i, F(x_t) + f(x_t; \phi))$$

gentleBoosting

At each iteration:
We chose $f_m(x)$ that minimizes the cost:

$$J(F + f_m) = \sum_{t=1}^{N} e^{-y_t(F(x_t) + f_m(x_t))}$$

Instead of doing exact optimization, gentle Boosting minimizes a Taylor approximation of the error:

$$J(F) \propto \sum_{t=1}^{N} e^{-y_t F(x_t)} (y_t - f_m(x_t))^2$$

At each iteration we just need to solve a weighted least squares problem

Weights at this iteration

Weak classifiers

• The input is a set of weighted training samples (x, y, w)

• Regression stumps: simple but commonly used in object detection.

\[ f_m(x) = a[x_k < \theta] + b[x_k \geq \theta] \]

Four parameters: \([a, b, \theta, k]\)

\[ b = E_w(y[x > \theta]) \]

\[ a = E_w(y[x < \theta]) \]
function classifier = gentleBoost(x, y, Nrounds)

... 

for m = 1:Nrounds

    fm = selectBestWeakClassifier(x, y, w);

    w = w .* exp(- y .* fm);

    % store parameters of fm in classifier

end

Initialize weights w = 1

Solve weighted least-squares

Re-weight training samples
Demo gentleBoosting

Demo using Gentle boost and stumps with hand selected 2D data:
> demoGentleBoost.m
Flavors of boosting

- AdaBoost (Freund and Shapire, 1995)
- Real AdaBoost (Friedman et al, 1998)
- LogitBoost (Friedman et al, 1998)
- Gentle AdaBoost (Friedman et al, 1998)
- BrownBoosting (Freund, 2000)
- FloatBoost (Li et al, 2002)
- …