

実世界情報実験 1

画像処理テーマ

補足：Adaboost

弱い識別器と強い識別器

- 弱い識別器 (2カテゴリに分類) $h_k(x)$

- なんでもよいが、偶然よりは少しだけ良い性能がある (誤り率が50%未満)

- 強い識別器

$$g(x) = \sum \alpha_k h_k(x)$$

- 弱い識別器の出力の重みづけ和
 - α をどのように決めるのが最適か？

AdaBoost

- すべての弱い識別器 h を一つずつ取り出し、学習サンプル集合に対して一番よい性能を出した h から順に選んで g に加えていく。
 - 現段の h をつかって識別をして誤ったサンプルの重みを増し、それ以外の重みを減らしたうえで次の h 選択を行う。
- ⇒ 苦手なサンプルにフォーカスした弱い識別器を選んでいく戦略

AdaBoost アルゴリズム

学習サンプル(データ、カテゴリ(1,-1)) $(x_1, y_1), \dots, (x_N, y_N)$

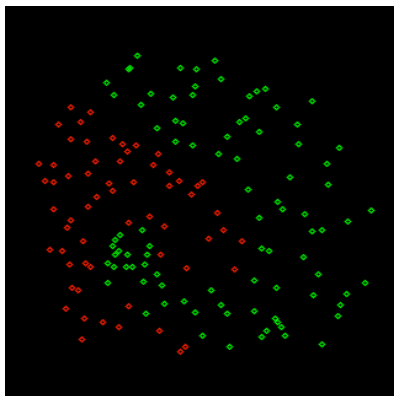
サンプルの重み $w_1(i) = 1/N \dots i = \{1, \dots, N\}$

1. initialize $g^{(0)}(x) = 0, k = 1$
2. すべての弱い識別器について、重み w_k に基づいて標本をとった学習サンプル集合を用いて弱い識別器 h_k を訓練する
3. $E_k \leftarrow h_k(x)$ の訓練誤差
4. $\alpha_k \leftarrow \frac{1}{2} \ln[(1 - E_k) / E_k]$
5. $w_{k+1}(i) \leftarrow w_k(i) / Z_k \cdot \begin{cases} e^{-\alpha_k} \dots h_k(x_i) = y_i \\ e^{\alpha_k} \dots h_k(x_i) \neq y_i \end{cases}$
6. $g^{(k)}(x) = g^{(k-1)}(x) + \alpha_k h_k(x)$
7. k が k_{\max} に達するか、誤り率が0なら終了
8. $k \leftarrow k + 1$ として2へ。

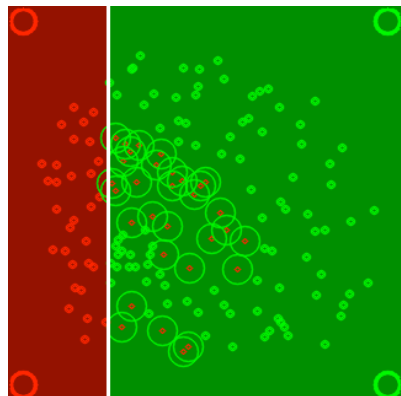
$$\sum_i e^{-y_i g(x_i)} \rightarrow \min$$

導出はWikipedia 英語版

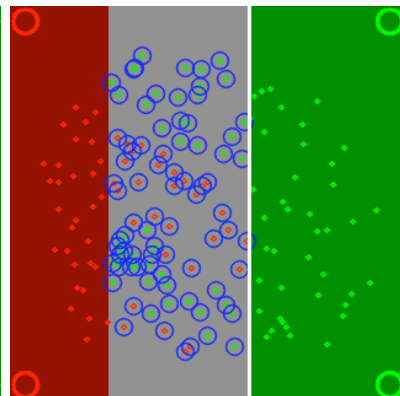
original



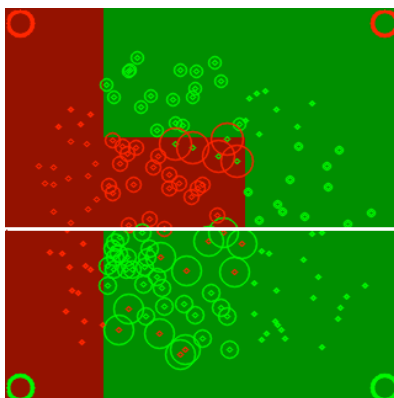
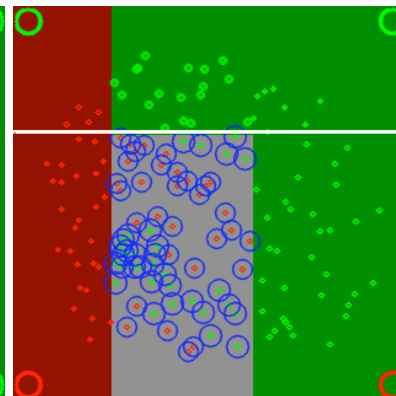
1st



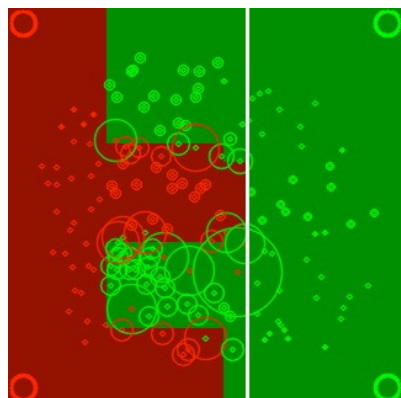
2nd



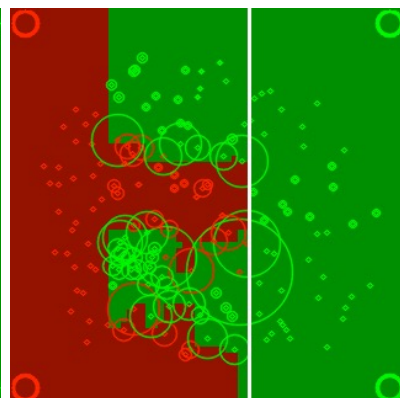
5th



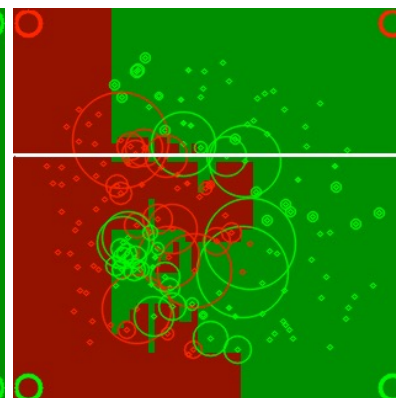
10th



20th

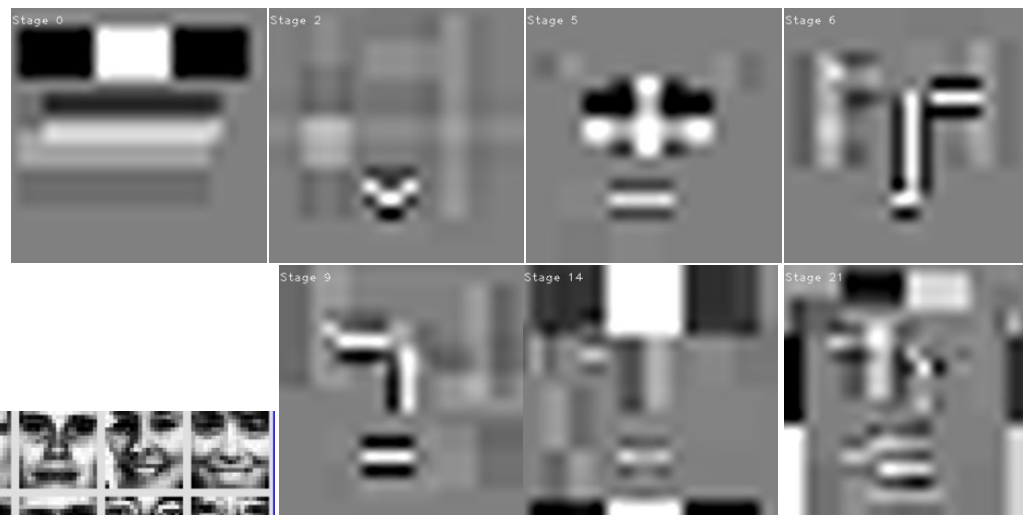


40th



50th

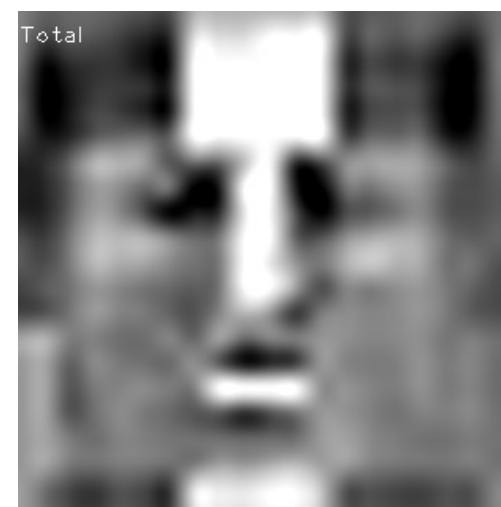
コンピュータから見た
人間の顔



これは顔



これは顔じゃない





コンピュータで顔さがし！

<http://demo.pittpatt.com/>