

# Coordinate Descent

[Breiman]

- $\{g_1, \dots, g_N\}$  = space of **all** weak classifiers
- then can write  $F(x) = \sum_t \alpha_t h_t(x) = \sum_{j=1}^N \lambda_j g_j(x)$
- want to find  $\lambda_1, \dots, \lambda_N$  to minimize

$$L(\lambda_1, \dots, \lambda_N) = \sum_i \exp \left( -y_i \sum_j \lambda_j g_j(x_i) \right)$$

- AdaBoost is actually doing **coordinate descent** on this optimization problem:
  - initially, all  $\lambda_j = 0$
  - each round: choose **one** coordinate  $\lambda_j$  (corresponding to  $h_t$ ) and update (increment by  $\alpha_t$ )
  - choose update causing **biggest decrease** in loss
- powerful technique for minimizing over huge space of functions