AdaBoost as Iterative Projection (cont.)

- algorithm:
  - start at $D_1 =$ uniform
  - for $t = 1, 2, \ldots$:
    - pick hyperplane/weak classifier $h_t \leftrightarrow g_j$
    - $D_{t+1} =$ (entropy) projection of D_t onto hyperplane
      $$= \arg \min_{D : \sum_i D(i) y_i g_j(x_i) = 0} \mathrm{RE}(D \parallel D_t)$$
- claim: equivalent to AdaBoost
- further: choosing h_t with minimum error $\equiv$ choosing farthest hyperplane