

AdaBoost is an Iterative-Projection Algorithm

[Kivinen & Warmuth]

- points = distributions D_t over training examples
- distance = relative entropy:

$$\text{RE}(P \parallel Q) = \sum_i P(i) \ln \left(\frac{P(i)}{Q(i)} \right)$$

- reference point \mathbf{x}_0 = uniform distribution
- hyperplanes defined by all possible weak classifiers g_j :

$$\sum_i D(i) y_i g_j(x_i) = 0 \Leftrightarrow \Pr_{i \sim D} [g_j(x_i) \neq y_i] = \frac{1}{2}$$

- intuition: looking for “hardest” distribution