Boosting (Mohri)

A concept class $\mathcal{C}$ is said to be weakly PAC-learnable if there exists an algorithm $\mathcal{A}$, $\gamma > 0$, and a polynomial function $poly(\cdot,\cdot,\cdot)$ such that for any $\delta > 0$, for all distributions D on $\mathcal{X}$ and for any target concept $c \in \mathcal{C}$, the following holds for any sample size $m \geq poly(1/\delta, n, size(c))$:

$$\mathbb{P}_{S \sim D^m}\left[R(h_S) \leq \frac{1}{2} - \gamma\right] \geq 1 - \delta,$$

where h_S is the hypothesis returned by algorithm $\mathcal{A}$ when trained on sample S. When such an algorithm $\mathcal{A}$ exists, it is called a weak learning algorithm for $\mathcal{C}$ or a weak learner. The hypotheses returned by a weak learning algorithm are called base classifiers.

ADABOOST$(S = ((x_1, y_1), \ldots, (x_m, y_m)))$

for $i \leftarrow 1$ to m do

  $D_1(i) \leftarrow \frac{1}{m}$ ← initially the distribution is uniform

for $t \leftarrow 1$ to T do ← rounds of boosting

  $h_t \leftarrow$ base classifier in $\mathcal{H}$ with small error $\epsilon_t = \mathbb{P}_{i \sim D_t}[h_t(x_i) \neq y_i]$

  $a_t \leftarrow \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t}$

  $Z_t \leftarrow 2[\epsilon_t(1 - \epsilon_t)]^{1/2}$ ▷ normalization factor (so that weights sum to 1)

  for $i \leftarrow 1$ to m do

    $D_{t+1}(i) \leftarrow (1/Z_t) \cdot D_t(i) \cdot \exp(-a_t y_i h_t(x_i))$

2020-2021 AdaBoost example

$S = \{(x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ $\quad y = \pm 1$ (labels)

$m = 5$, οπότε $D_1(i) = 1/5$, $i = 0, \dots, 4$

$\epsilon_t = 0.4$, $A = 54$, $(100 + A) \bmod 5 = 4$, $(54 - 99) \bmod 5 = 0$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$(x_4, y_4) \qquad\qquad (x_0, y_0)$$

$a_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} = \frac{1}{2} \ln \frac{3}{2}$, $Z_t = 2\left[\epsilon_t (1 - \epsilon_t)\right]^{1/2} = 2\sqrt{0.6 \cdot 0.4} = 0.4\sqrt{6}$

Ισχύει $y_i h_1(x_i) = \begin{cases} +1, & \text{εάν } i = 1, 2, 3 \\ -1, & \text{εάν } i = 0, 4 \end{cases}$

Άρα,

- $D_2(0) = \frac{1}{0.4\sqrt{6}} \cdot \frac{1}{5} \cdot e^{\ln \sqrt{\frac{3}{2}}} = \frac{1}{2\sqrt{6}} \cdot \sqrt{\frac{3}{2}} = \frac{1}{4}$

- $D_2(1) = \frac{1}{0.4\sqrt{6}} \cdot \frac{1}{5} \cdot e^{-\ln \sqrt{\frac{3}{2}}} = \frac{1}{2\sqrt{6}} \cdot \sqrt{\frac{2}{3}} = \frac{1}{6}$

αντίστοιχα, $D_2(2) = D_2(3) = \frac{1}{6}$ και $D_2(4) = \frac{1}{4}$

Sanity check:

- D_2 (misclassified) $>$ D_2 (correctly classified) ✓
- $\sum_{i=0}^{4} D_2(i) = \frac{1}{4} \cdot 2 + \frac{1}{6} \cdot 3 = 1$ ✓