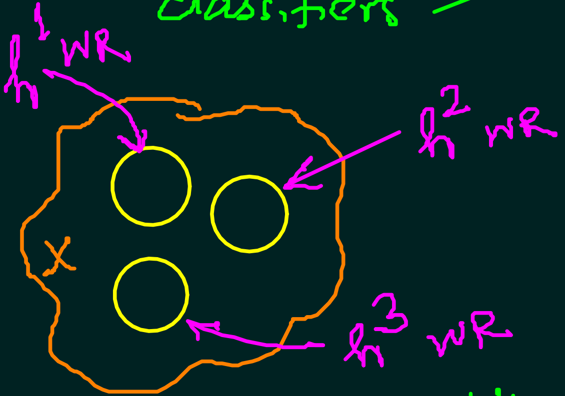
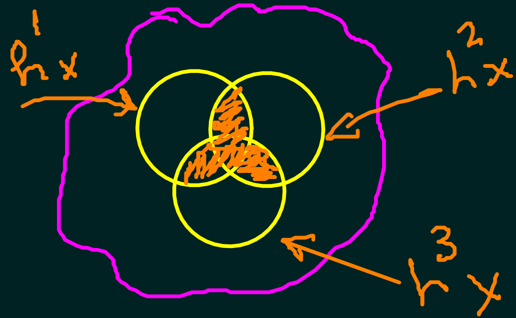
Binary classifier, $h : \mathcal{X} \rightarrow [+1, -1]$ $\in$ [0.0 (Strong) —— 0.5 (weak) —— 1.0]

Weak classifiers $\Rightarrow$ vote $\rightarrow$ $\mathcal{H}$ (Strong Classifier) $\leftarrow$ Ensemble Learning





$$\mathcal{H}(x) = \text{sign}\left(h^1(x) + h^2(x) + h^3(x)\right)$$

w_i $\qquad$ w_N

DATA $\langle x_1, y_1 \rangle$ ----- $\langle x_N, y_N \rangle$

$$\boxed{W_i^{(1)}} = \frac{1}{N}$$

$\rightarrow$ $\mathcal{D}_1$ $\rightarrow$ h^1 $\quad \alpha^1$

$\rightarrow$ Data with Exaggeration (h^1 is wrong $w_i \uparrow$) $\rightarrow$ $\mathcal{D}_2$ $\rightarrow$ h^2 $\quad \alpha^2$

$\rightarrow$ Data with exaggeration (h^1 & h^2 differs) $\rightarrow$ $\mathcal{D}_3$ $\rightarrow$ h^3 $\quad \alpha^3$

$\rightarrow$ ....

$$\boxed{\mathcal{H}(x)} = \text{sign}\left[\alpha^1 h^1(x) + \alpha^2 h^2(x) + \alpha^3 h^3(x) + \cdots\right]$$

(Major) Weighted wisdom of expert crowds

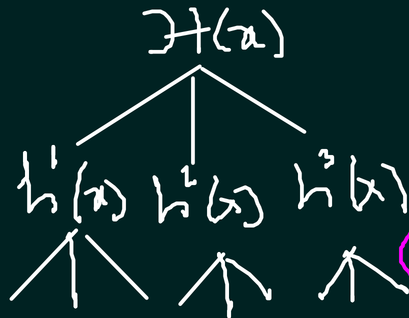
$w_i^{(t)}$

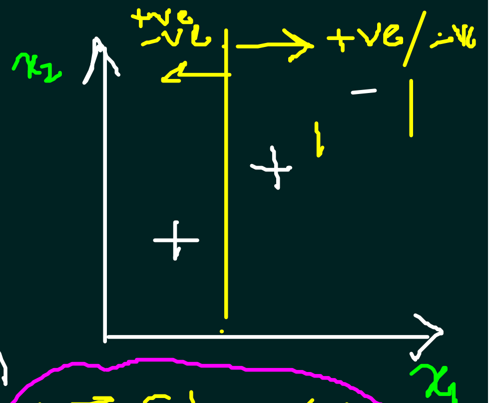
$\alpha^{(t)}$

$\boxed{\text{AdaBoost}}$ 1997

$$\mathcal{H}(x) = \text{sign}\left[\alpha^1 h^1(x) + \alpha^2 h^2(x) + \cdots\cdots\right]$$

$$\mathcal{H}^t(x) = \text{sign}\left[\mathcal{H}^{t-1}(x) + \underbrace{\alpha^t h^t(x)}\right]$$





DT Stumps

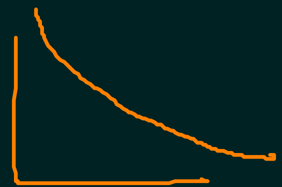
$$W_i^{(1)} = \frac{1}{N} \quad \forall\, i \in [1, N]$$

+1
-1

$$\epsilon^{(1)} = \sum_{\text{WRONG}} W_i^{(1)} \to \frac{1}{N} \longrightarrow \epsilon^{(t)} = \sum_{h^t(x) \neq y} W_i^{(t)}$$

$$\sum_{i=1}^{N} W_i^{(t)} \quad 1$$

h^t $\quad$ + ✓ $\quad$ + ✓ $\quad$ $\boxed{+ \times}$ $\quad$ $\boxed{- \times}$ $\quad$ $\boxed{- \times}$ $\quad$ − ✓ $\quad$ − ✓ $\quad$ $\boxed{+ \times}$ $\quad$ + ✓



$$\left[\text{assume} \quad \sum_{i=1}^{N} W_i^{(t)} = 1\right]$$

$\langle x, y \rangle$

① $\boxed{W_i^1 = \frac{1}{N}}$ $\quad$ AdaBoost

② $\boxed{\text{Pick } h^t \text{ that min. } \epsilon^t} \longrightarrow \boxed{\text{Pick } \alpha^t}$?

$\boxed{\text{Update } W_i^{t+1}}$ ✓

▷ exp. loss : $\quad e^{-\alpha^t \underbrace{h^t(x)\, y(x)}}$

±1 $\quad$ ±1

$$W_i^{t+1} \leftarrow \frac{W_i^t}{Z} e$$

WR → −ve

COR → +ve

$$\frac{W_i^t}{Z} < \frac{W_i^t}{Z} e^{\alpha^t}$$

$$\frac{W_i^t}{Z} \quad \frac{1}{e^{\alpha^t}}$$

$$E^t = \sum w_i^t \, e^{-\underline{\alpha^t}\, \underline{h^t(x)\, y(x)}} \qquad \frac{\partial E}{\partial \alpha^t} = 0$$

$$E^t = \sum_{WR} w_i^t \, e^{\alpha^t} + \sum_{COR} w_i^t \, e^{-\alpha^t} \Rightarrow e^{\alpha^t} \sum_{WR} w_i^t - e^{-\alpha^t} \sum_{COR} w_i^t = 0$$

$$\alpha^t + \ln\left(\sum_{W} w_i^t\right) = -\alpha^t + \ln\left(\sum_{COR} w_i^t\right) \qquad \boxed{\epsilon^t = \sum_{W} w_i^t}$$

$$\Rightarrow \alpha^t = \frac{1}{2} \ln\left(\frac{\sum_{C} w_i^t}{\sum_{W} w_i^t}\right) = \frac{1}{2} \ln\left(\frac{1-\epsilon^t}{\epsilon^t}\right)$$

Pick $\boxed{\underline{\alpha^t} = \frac{1}{2} \ln\left(\frac{1-\epsilon^t}{\epsilon^t}\right)}$

$$H^t(x) = \text{sign}\left[\underline{H^{t-1}(x)} + \underline{\alpha^t}\, h^t(x)\right]$$

$w_i^t \Rightarrow w_i^{t+1} \Big\} \longrightarrow \alpha^{t+1}$ pick up

① $W_i^{t+1} \leftarrow \dfrac{W_i^t}{Z} e^{-\alpha^t h^t(x) \cdot y(x)}$

② $\alpha^t \leftarrow \dfrac{1}{2} \ln\left(\dfrac{1-\epsilon^t}{\epsilon^t}\right)$

$$\epsilon^t = \sum_{WR} W_i^t$$

$$W_i^{t+1} = \frac{W_i^t}{Z} \begin{cases} \sqrt{\dfrac{\epsilon^t}{1-\epsilon^t}}, & \text{CORRCT} \\ \sqrt{\dfrac{1-\epsilon^t}{\epsilon^t}}, & \text{WRONG} \end{cases} \Bigg\} \; \alpha$$

$$\boxed{\sum_{i=1}^{N} W_i^{t+1} = 1} \longrightarrow \sum_{COR} \frac{W_i^t}{Z} \sqrt{\frac{\epsilon^t}{1-\epsilon^t}} + \sum_{WRO} \frac{W_i^t}{Z} \sqrt{\frac{1-\epsilon^t}{\epsilon^t}} = 1$$

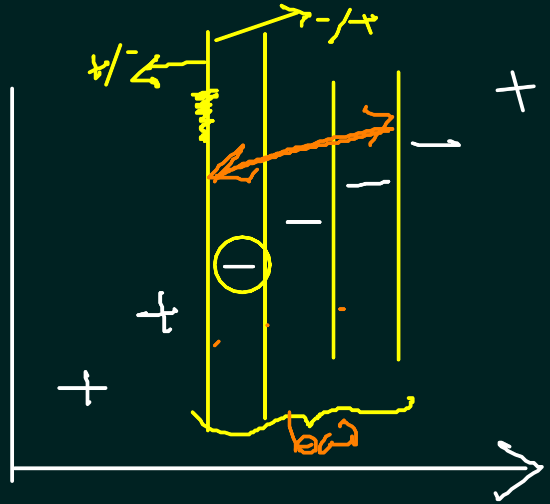
$$\Rightarrow \sqrt{\frac{\epsilon^t}{1-\epsilon^t}} \underbrace{\sum_{COR} W_i^t}_{(1-\epsilon^t)} + \sqrt{\frac{1-\epsilon^t}{\epsilon^t}} \underbrace{\sum_{WR} W_i^t}_{\epsilon^t} = Z$$

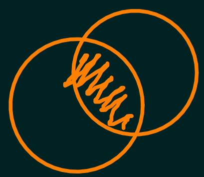
$$\frac{W_i^t}{2(1-\epsilon^t)}$$

$$\Rightarrow Z = \boxed{2\sqrt{\epsilon^t(1-\epsilon^t)}}$$

H_1

$$\left[\sum_{COR} W_i^{t+1} = \frac{1}{2} \;\middle|\; \sum_{W} W_i^{t+1} = \frac{1}{2} \right]$$

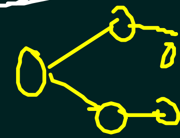




h^1 —
h^2 —
h^3 —
⋮
} Indep.

h^1

Y. Freud
R. Schapire
1997
1999



$T << dN$ stumps → hope-2

$$\boxed{H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha^t h^t(x)\right)}$$

1. It converges faster → exp
2. This does not overfit → Practical !!
3. Computationally faster

} AdaBoost Algo

↓

Netflix

$(X_1, Y_1)^{1/N}$ - - - - $(X_N, Y_N)^{1/N}$

→ D_1 → h^1 → ✓ ✓ ✗

Boost

D_2 → h^2 . . . . $h^T(x)$

↓ w_i^{t+1}

corr. $\boxed{\dfrac{w_i^t}{Z} e^{-\alpha^t}}$

→ ↑ w_i^{t+1} ← $\dfrac{w_i^t}{Z} e^{\alpha^t}$

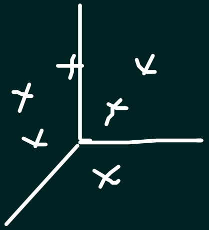
$$\alpha^t = \frac{1}{2} \ln\left(\frac{1-\epsilon^t}{\epsilon^t}\right)$$

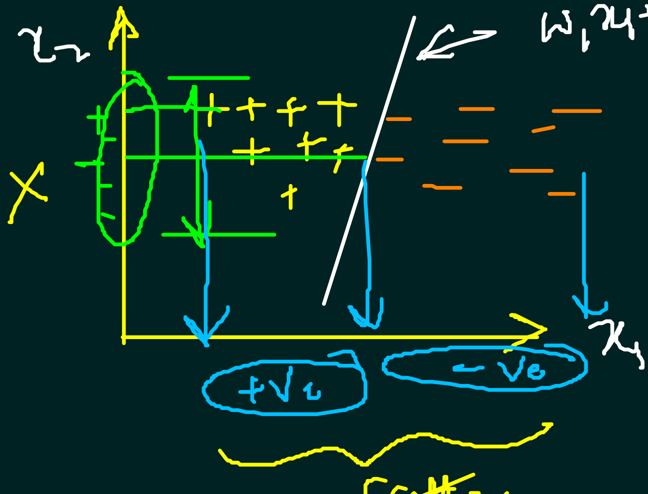
Dimensionality Reduction:

$(d^k) \leftarrow$ computation

10000

{ Reg. ✓
AdaBoost ✓
Dim. ↓





$\{x_1 \cdots x_d\}$

$\longrightarrow \underbrace{\{x_{k_1} \cdots x_{k_m}\}}$ → +ve
→ -ve

→ $\boxed{\text{Kernel}}$ → ↑ Dimen

$\boxed{\infty}$

2

↳ Metric

↳ How attributes chosen?

↳ PCA (Principal Component Analysis)

1