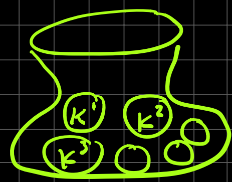
Ada Boost

Adaptive Boosting

We have many "experts" (opinions between -1 and 1)



← classifiers

We want to put together a team of classifiers:

$$C(\vec{x}_i) = \alpha_1 k_1(\vec{x}_i) + \alpha_2 k_2(\vec{x}_2) + \cdots + \alpha_\ell k_\ell(\vec{x}_i)$$

↑ weigths ↖ classifiers

The classifiers are given, we want to collect their oppinions and the classificaton is then

$$\text{sign}(C(\vec{x}_i))$$

Since the classifiers are fixed we test all of them once with the complete data set:

Classifiers

Data	1	2	3	....	L
$\vec{x}_1$	0	1	0		1
$\vec{x}_2$	0	0	1		0

$\vdots$

$\vec{x}_N$ | 1 1 0 0 ← hit (correct)

↑ miss (incorrect)

Drafting

$$\operatorname{sgn}[C(x)] = \overset{\operatorname{sgn}}{[}\alpha_1 k_1(x) + \alpha_2 k_2(x)]$$

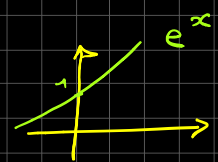
	α_1	α_2	...	Loss function
Data ↓	k_1	k_2	....	
x_1	$e^{-\alpha_1 y_1 k_1(x_1)}$	$e^{-\alpha_2 y_1 k_2(x_1)}$		→ Mult +
x_2	$e^{-\alpha_1 y_2 k_1(x_2)}$	$e^{-\alpha_2 y_2 k_2(x_2)}$		→ Mult +
$\vdots$	$\vdots$			$\vdots$
x_N	$e^{-\alpha_1 y_N k_1(x_N)}$	$e^{-\alpha_2 y_N k_1(x_N)}$		→ Mult +
				Loss

If hit $\quad e^{-\alpha_1 y_1 k_1(x_1)} = e^{-\alpha_1}$

If miss $\quad = e^{+\alpha_1}$

$$e^{+\alpha_1} > e^{-\alpha_1}$$

(weigths are positive)



If we want to bring a new expert in :

	α_1	α_2	...	α_{m-1}		new α_m
	k_1	k_2		k_{m-1}		k_m
x_1					→ $\omega_1^{(m)}$	$e^{-\alpha_m y_1 k_m(x_1)}$ +
x_2					→ $\omega_2^{(m)}$	$e^{-\alpha_m y_2 k_m(x_2)}$ +
$\vdots$		$\vdots$				$\vdots$ +
x_N					→ $\omega_N^{(m)}$	+
						Loss

$$\text{Loss} = \sum^{N} \omega_i^{(m)} e^{-\alpha_m y_i k_m(x_i)}$$

where $$\omega_i^{(m)} = e^{-y_i C_{m-1}(x_i)}$$

committe of m-1 experts

The $\omega_i^{(m)}$ can be understood as weigths for each data point

We can express the loss as:

$$E = \sum_{\text{hits}} \omega_i^{(m)} e^{-\alpha_m} + \sum_{\text{misses}} \omega_i^{(m)} e^{\alpha_m}$$

$$\boxed{\text{Min} \quad E = W_h e^{-\alpha_m} + W_e e^{\alpha_m}}$$

1 $$\text{Min} \quad e^{\alpha_m} E = W_c + W_e e^{2\alpha_m}$$

$$= \underbrace{(W_c + W_e)}_{\substack{\text{total loss} \\ \text{of classifier with} \\ m-1 \text{ experts}}} + W_e \underbrace{(e^{2\alpha_m} - 1)}_{\text{positive}}$$

Minimize E $\Rightarrow$ pick lowest W_e

i.e. lowest rate of weighted error

2 Weigths

$$\frac{dE}{d\alpha_m} = -W_h e^{-\alpha_m} + W_e e^{\alpha_m}$$

$$\Rightarrow \quad -W_h + W_e e^{2\alpha_m} = 0$$

$$\Rightarrow \quad e^{2\alpha_m} = W_h / W_e$$

$$\Rightarrow \quad \boxed{\alpha_m = \frac{1}{2} \ln\left(\frac{W_h}{W_e}\right)}$$

$W = W_h + W_e$

or

$$\alpha_m = \frac{1}{2} \ln\left(\frac{W - W_e}{W_e}\right) = \frac{1}{2} \ln\left(\frac{1 - e_m}{e_m}\right)$$

where $e_m = W_e / W$

(percentage of weighted error in total weigth)

Ada Boost

Initialize

$w_i^{(1)} = 1 \quad \forall i$

mth step

1) Select from the pool of classifiers, k_m which minimizes

$$W_e = \sum_{\text{misses}} w_i^{(m)}$$

2) $\alpha_m = \frac{1}{2} \ln\left(\frac{1 - e_m}{e_m}\right)$, $e_m = W_e / W$

3) Update weigths:

$$w_i^{(m+1)} = w_i^{(m)} \sqrt{\frac{1 - e_m}{e_m}} \quad \text{for miss}$$

$$w_i^{(m+1)} = w_i^{(m)} \sqrt{\frac{e_m}{1 - e_m}} \quad \text{for hit}$$

Boosted Trees

→ AdaBoost using a bag of trees

<u>SVM's with boosting</u>

$$c(\vec{x}) = \sum \underset{\uparrow \text{ weigths}}{\alpha_i} \underbrace{y_i K(x_i, \vec{x})}_{\text{individual classifiers}}$$

Epiphany

$$\text{Min } E = W_h e^{-\alpha_m} + W_e e^{+\alpha_m}$$

$$\alpha_m = \frac{1}{2} \ln\left(\frac{W_h}{W_e}\right) = \ln\left(\sqrt{\frac{W_h}{W_e}}\right)$$

$$E = W_h \sqrt{\frac{W_e}{W_h}} + W_e \sqrt{\frac{W_h}{W_e}}$$

$$\text{Min } E = 2\sqrt{W_h W_e}$$

$$\text{Min } W_h W_e = \text{Min } (W - W_e) W_e$$

$$= \text{Min } (\underbrace{W \cdot W_e - W_e^2}_{})$$

Pool of classifiers
4998

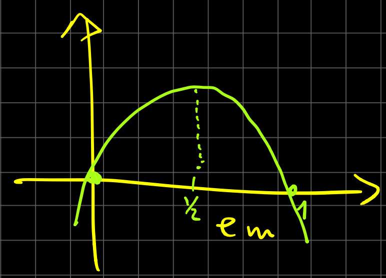
select the next classifier
with lowest $\boxed{W_e}$

Another way of seeing it:

$$\text{Min } (W \cdot W_e - W_e^2) = \text{Min } \frac{1}{} (W W_e - W_e^2)$$

$$= \operatorname{Min} \left(e_m - e_m^2 \right)$$

$$0 \leq e_m \leq 1$$



the lower e_m, the better

If $e_m > \frac{1}{2}$, flip the classifier