5444NA
9/24/2013

LINEAR CLASSIFIER : $a_k^n = \vec{w}^T \vec{\phi}^n$

TRAINING CRITERIA

① LINEAR MSE $\mathcal{E} = \sum_{n=1}^{N} |t^n - a^n|^2$

$\Rightarrow W^T = \Phi^\dagger T = (\Phi^T \Phi)^{-1} \Phi^T T$

$\vec{w} = (\Phi^T \Phi)^{-1} \Phi^T \vec{t}$

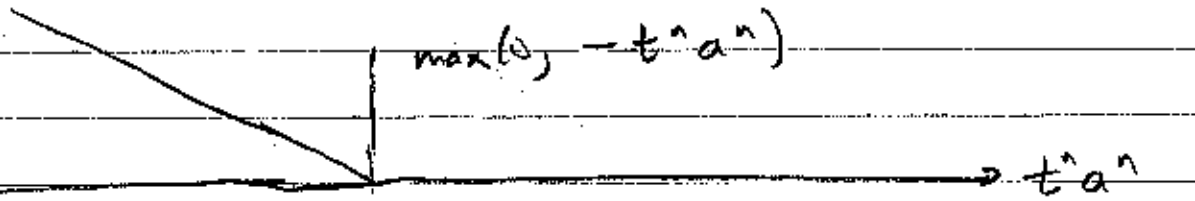
② NONLINEAR MSE $\mathcal{E} = \sum_{n=1}^{N} |t^n - h(a^n)|^2$

$w_j = w_j - \eta \frac{\partial \mathcal{E}}{\partial w_j}$

$= w_j + \eta (t_i^n - h(a^n)) h'(a_i^n) \phi_j^n$

③ PERCEPTRON $\mathcal{E} = \sum_{n=1}^{N} \max(0, -t_i^n a_i^n)$

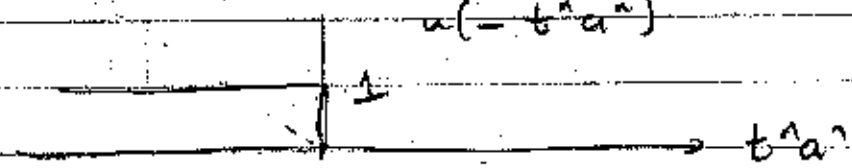
$t_i^n a^n \begin{cases} > 0 & \text{IF } t^n, a^n \text{ SAME SIGN} \\ < 0 & \text{IF OPPOSITE SIGN} \end{cases}$



TOKEN $\vec{\phi}^n$ IS MISCLASSIFIED

IFF $\text{sign}(t^n) \neq \text{sign}(a^n)$ ← UNIT STEP

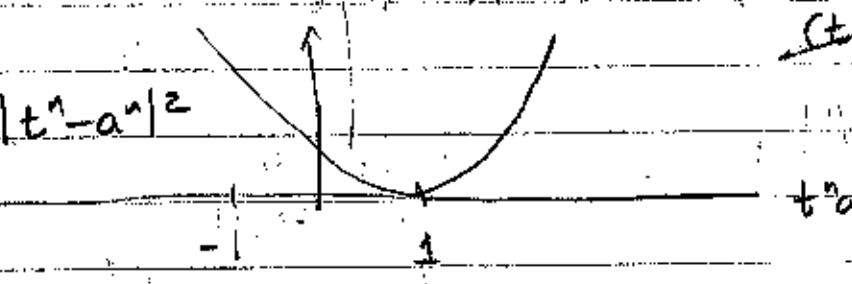
$\Leftrightarrow$ (# ERRORS) $= \sum_{n=1}^{N} u(-t^n a^n)$



LINEAR MSE: IF $t^n \in \{-1, 1\}$, THEN

$|t^n - a^n|^2 = 1 - 2t^n a^n + (a^n)^2$

$= 1 + 2t^n a^n + \frac{(t^n a^n)^2}{(t^n)^2}$ → 1



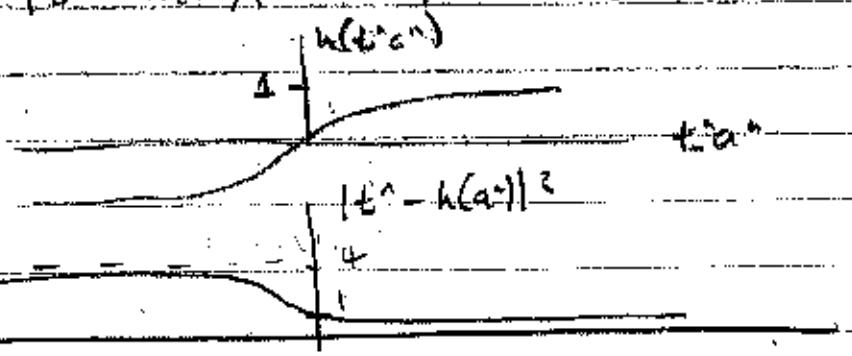
NONLINEAR MSE: $|t^n - h(a^n)|^2$

$= 1 - 2t^n h(a^n) + h^2(a^n)$

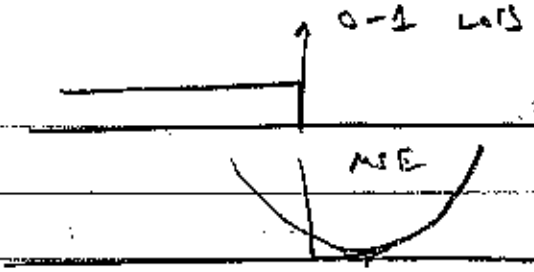
BUT $h(-a^n) = -h(a^n)$

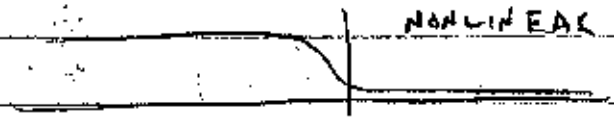
$h(t^n a^n) = t^n h(a^n)$

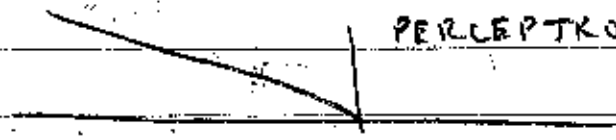
$|t^n - h(a^n)|^2 = 1 - 2h(t^n a^n) + h^2(t^n a^n)$



LOSS FUNCTIONS







	ADVANTAGES	DISADVANTAGES
0-1 LOSS	IT'S WHAT YOU REALLY CARE ABOUT	NOT CONVEX ⇒ NO GLOBAL OPTIMUM NOT DIFFERENTIABLE ⇒ NO GRADIENT SOLUTION EITHER!
LINEAR MSE	CONVEX ⇒ HAS GLOBAL OPTIMUM	VERY GOOD TOKENS AS BAD AS VERY BAD TOKENS
NONLINEAR MSE	CLOSE TO 0-1 LOSS; DIFFERENTIABLE	"VERY BAD TOKENS" HAVE $h'(a_n^k) \approx 0$ ⇒ GRADIENT DESCENT "GIVES UP"
PERCEPTRON	FIXES THIS PROBLEM; GRADIENT DESCENT IS VERY EASY!!	

GRADIENT DESCENT :

$$E = \sum_{n \in \mathcal{T}} \max(0, -t^n a^n)$$

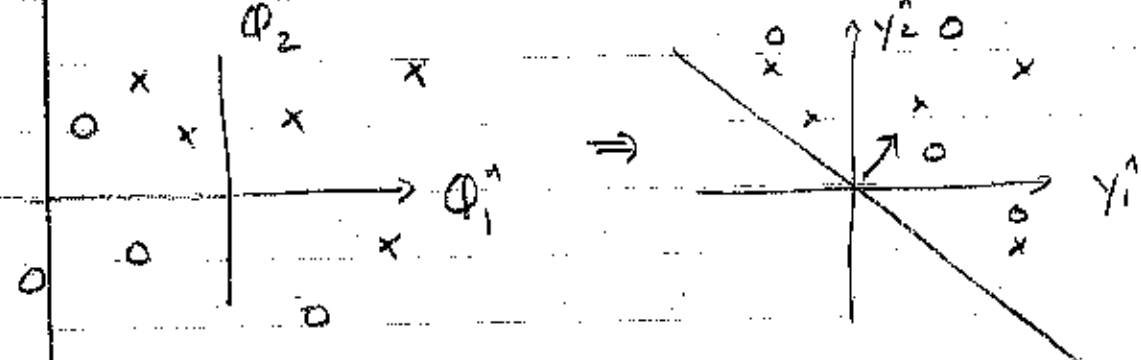
$$= \sum_{n=1}^{N} \max\left(0, -t^n \sum_{j=1}^{M} w_j \phi_j^n\right)$$

$$\frac{\partial}{\partial w_j} \max(0, -t^n a^n) = \begin{cases} 0 & t^n a^n \geq 0 \\ -t^n \phi_j^n & t^n a^n < 0 \end{cases}$$

$$\nabla_{\vec{w}} E = -\sum_{n:\ t^n a^n < 0} t^n \vec{\phi}^n$$

GEOMETRIC INTERPRETATION

$$\text{LET } \vec{y}^n = t^n \vec{\phi}^n = \begin{cases} \vec{\phi}^n & \text{IF } t^n = +1 \\ -\vec{\phi}^n & \text{IF } t^n = -1 \end{cases}$$



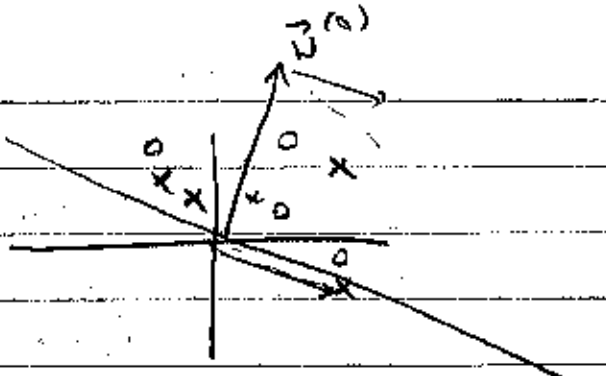
o: $t^n = -1$ ZERO ERROR:

x: $t^n = +1$ FIND $\vec{w}$ SUCH THAT

$$\vec{w}^T \vec{y}^n > 0 \quad \text{FOR ALL } n$$

PERCEPTRON :

INITIALIZE: $\vec{w} = \sum \vec{y}^{\wedge}$



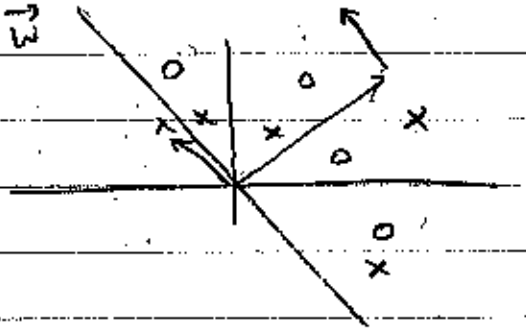
ITERATE

ⓐ FIND THE $\vec{y}^{\wedge}$ S.T. $\vec{w}^T \vec{y}^{\wedge} < 0$

ⓑ ADD THEM TO $\vec{w}$

$\vec{w} = \vec{w} + \sum \vec{y}^{\wedge}$

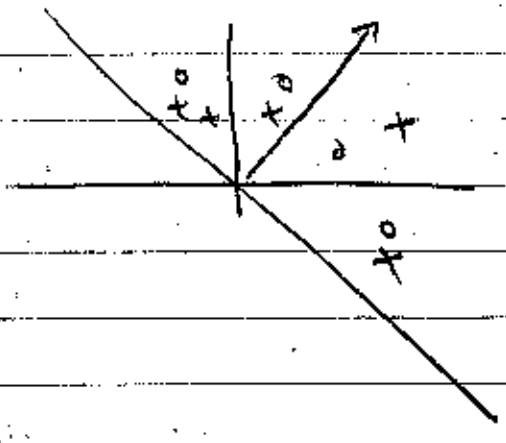
$\vec{w}^T \vec{y}^{\wedge} < 0$



TERMINATE

WHEN ALL $\vec{y}^{\wedge}$

CORRECTLY CLASSIFIED



CONVERGENCE :

- $\vec{w}$ KEEPS GETTING BIGGER
- PROCESS TERMINATES IF DATA ARE LINEARLY SEPARABLE
- IF NOT, SET A THRESHOLD W_{MAX}; STOP WHEN $\|\vec{w}\|$ PASSES W_{MAX}

PROOF OF CONVERGENCE

LINEARLY SEPARABLE DATA

$\Leftrightarrow$ $\exists\ \hat{w}$ SUCH THAT $\text{sign}(\hat{w}^T \vec{y}^n) = t^n$

$\Leftrightarrow$ $\hat{w}^T \vec{y}^n > 0$ FOR ALL n,

$\|\hat{w}\| = 1$

BUT NOTICE, AT τ^{TH} ITERATION,

$$\vec{w}_{(\tau+1)} = \vec{w}_{(\tau)} + \sum_{n:\, \vec{w}_{(\tau)}^T \vec{y}^n < 0} \vec{y}^n \qquad (< 0 !!)$$

$$\|\vec{w}_{(\tau+1)}\|^2 = \|\vec{w}_{(\tau)}\|^2 + \Big\|\sum_{w^T y < 0} \vec{y}^n\Big\|^2 + 2\vec{w}_{(\tau)}^T \sum_{w_{(\tau)}^T y < 0} \vec{y}^n$$

$$\leq \|w_{(\tau)}\|^2 + \|\vec{s}\|^2$$

SO $\boxed{\|w_{(\tau)}\| \leq \sqrt{\tau}\ \|\vec{s}\|, \quad \vec{s} = \sum_n \vec{y}^n}$

τ = # ITERATIONS THAT HAVE ANY ERRORS

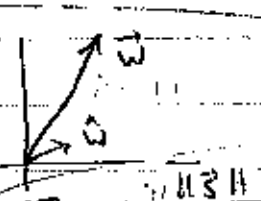
BUT NOTICE $\vec{w}_{(\tau)} = \sum_{n=1}^{N} M^n \vec{y}^n$

M^n = # TIMES $\vec{y}^n$ HAS BEEN MISCLASSIFIED

$$\boxed{\hat{w}^T \vec{w}_{(\tau)} = \sum_{n=1}^{N} M^n (\hat{w}^T \vec{y}^n) \geq N\tau \min_n (\hat{w}^T \vec{y}^n)}$$

BUT $\hat{w}^T \vec{w}_{(\tau)} = \|\vec{w}_{(\tau)}\| \cos\theta$

$\leq \|\vec{w}_{(\tau)}\|$



SO $N\tau \min_n (\hat{w}^T \vec{y}^n) \leq \sqrt{\tau}\ \|\vec{s}\|$: $\boxed{\sqrt{\tau} \leq \dfrac{\|\vec{s}\|}{N \min_n (\hat{w}^T \vec{y}^n)}}$