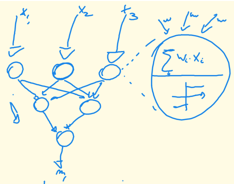


σ: ACT. FUNC.

W_{ij}: weights

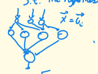
$a_{t,i}$: Result Lin. Combination

$O_{t,i}$: $\sigma(a_{t,i})$

boolean F.

$\{+1,-1\}^m \to \{+1,-1\}$

Claim: For every m there is a network of depth 2.
S.t. the hypothesis class contains all m-Boolean func.



$\vec{x} = \vec{u}_i$

For every vector u_i for which $f(u_i)$ is TRUE
We add a neuron in the hidden layer that checks if $x = u_i$
i.e. the neuron implements the function

$$g(x) = \text{sign}(\langle x, u \rangle - m + 1)$$

$$f(x) = \text{sign}\left(\left(\sum_i g_i(x)\right) + k - 1\right.$$

then

$$|V| = O(2^m)$$

Theorem: For every m, let $s(m)$ be the minimal integer s.t. $\exists$ graph (V,E) with $|V| = s(m)$ and whose hypothesis class contains all m-Boolean Functions. Then $s(m)$ is exponential in m.

Assumption: $VC(H_{V,E,\text{sign}}) = O(|V|^3)$

$VC\dim(H_{V,E,\text{sign}}) = 2^m$

$$|V| \geq \Omega(2^{m/3})$$

$[-1,1]^m \to \{-1,+1\}$

Theorem: For $T: \mathbb{N} \to \mathbb{N}$, $\forall m \in \mathbb{N}$,
F_m is the set of functions that can be implemented with a Turing Machine in time $O(T(m))$. Then, $\exists\, b,c \in \mathbb{R}$ s.t. $\forall m$ there exists a graph (V,E) with $|V| \leq cT^2 + b$ and whose hypothesis class contains F_m

$$\Pi(m) = \max_{C \subseteq X, |C| = m} |H_C| \leq 2^m$$

growth function:
the maximal number $h \in H$ that shatter a set of m points. $\Pi_H(C)$. $|C| = m$
ASSUMPTION: growth f. of a composition of hypothesis classes is bounded by the product of the growth function value for each hypothesis class.

Theorem: VC dimension of $H_{V,E,\text{sign}}$ is $O(|E|\log|E|)$

$$H_{V,E,\text{sign}} = H_T \circ H_{T-1} \cdots \circ H_1$$

ACC. TO ASSUMPTION:

$$\tau_H(m) \leq \prod_{t=1}^{T} \tau_{H^t}(m) \leq \prod_{i=1}^{|V_t|} \tau_{H^{t,i}}(m)$$

VC dim of halfspace is d.

by Sauer's theorem

$$\tau_{H^{t,i}} \leq \left(\frac{em}{d_{t,i}}\right)^{d_{t,i}} \leq (em)^{d_{t,i}}$$

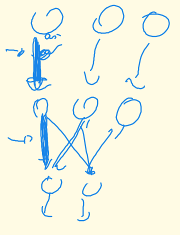
$$\tau_H(m) \leq (em)^{\sum_{t,i} d_{t,i}} \leq em^{|E|}$$

IF there are m shattered points, then,

$\tau_H(m) = 2^m$ and

$$2^m \leq (em)^{|E|} \Rightarrow m \geq |E|\log(em)/\log(2)$$

$$\nabla_x f(g(x))$$



$$\frac{\partial E}{\partial W_{t=T,j}} = \frac{\partial E}{\partial O_{t=T,i}} \cdot \frac{\partial \sigma(a)}{\partial a_{t=T,i}} \cdot \frac{\partial a}{\partial W_{t=T}}$$

$$\underbrace{(O_{t=T,i} - y_i)}_{\delta} \cdot \sigma'(a) \cdot O_{t=T-1,j}$$

$$\frac{\partial E}{\partial W_{t=T-1,ij}} = \underbrace{\frac{\partial E}{\partial O_{t=T-1,i}}}_{\delta} \cdot \frac{\partial O_{t=T-1,i}}{\partial a_{t=T-1,i}} \cdot \frac{\partial a}{\partial W_{t=T-1,i}}$$

$$\sigma'(O_{t=T-1}) \cdot O_{t=T-2}$$

$$\frac{\partial E}{\partial O_{t-1,i}} = \sum_{j=1}^{k_{t=T}} \frac{\partial E}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial a_j} \cdot \frac{\partial a}{\partial O_{t-1}}$$

$$\delta_j \quad \sigma'(a) \cdot W_{ij}$$