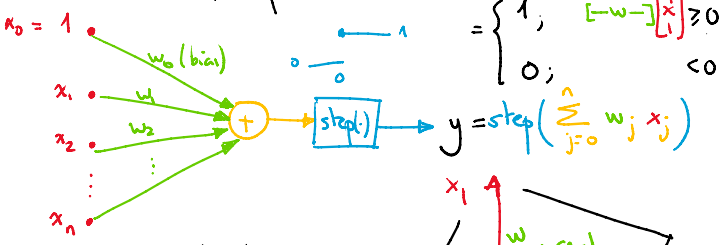
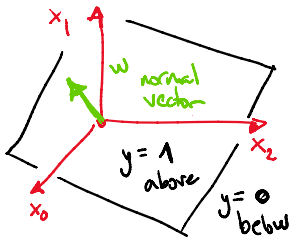
ECE/CS 559 Lecture 5 9/10

Last time: Approximation Theorems and Learning Algorithms
Perceptrons

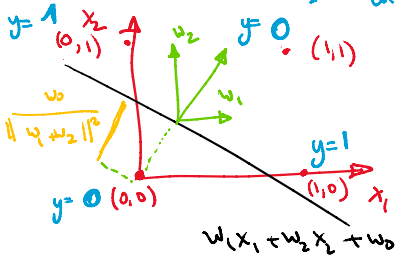


- What can it do?
  - linear separator
  - Can model logic gate



- How many layers are sufficient for any operation?
  Answer: Two. Encode the truth table (CNF/DNF Theorem)
  May not be the most efficient - Depth ⇒ fewer neurons
- Is 1 layer sufficient? Answer: No!

Example: XOR gate:



cannot create XOR pattern with a linear separator (without x_0, does not pass through origin) affine

① Example

$$\underbrace{\neg x_1 \wedge \neg x_2 \wedge \cdots \wedge \neg x_n}_{n \text{ negated}} \wedge \underbrace{x_{n+1} \wedge \cdots \wedge x_{n+m}}_{m \text{ direct}} \quad \text{(AND)}$$

$$\Leftrightarrow \quad (1-x_1) + \cdots + (1-x_n) + x_{n+1} + \cdots + x_{n+m} \geq n+m-\tfrac{1}{2}$$

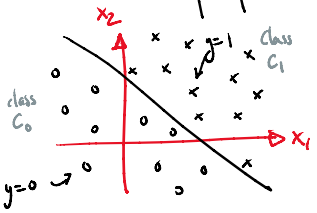
satisfied if all $= 1$

$$\Leftrightarrow \quad \underbrace{-n+\tfrac{1}{2}}_{w_0} \underbrace{-x_1}_{w_1} \underbrace{-x_2}_{w_2} - \cdots \underbrace{-x_n}_{w_n} + \underbrace{x_{n+1}}_{w_{n+1}} + \cdots + \underbrace{x_{n+m}}_{w_{n+m}} \geq 0$$

- Because NOT + either AND or OR are universal Multilayer Perceptrons can represent any logical operation.

② Perceptron Learning Algorithm

- Given a description of the desired behavior $y = f(x)$ how do we find parameters such that $y_{\text{params}}(x) \approx f(x)$? (in this case params $\equiv w$)
- Assume there exists w such that $y_w(x) = \text{step}(w^T x) = y$ exactly.
  - Jargon: Classes ($y=0$ $y=1$ regions) are called linearly separable since perceptron can realize this, the problem is realizable



- In supervised learning, the description takes the form of data: many pairs (x, y)

Algorithm: Input: Data $= \{(x,y), \cdots\}$ $x \in \mathbb{R}^n$ $y \in \{0,1\}$, η
Output: weights $w \in \mathbb{R}^n$

1. Initialize w arbitrarily.
2. While there exists $(x,y) \in$ Data such that $y_w(x) \neq y$:
3. For each $(x,y) \in$ Data:
   - If $y_w(x) = y$: do nothing
   - Else if $y_w(x) = 0$ and $y = 1$: $w \leftarrow w + \eta x$
   - Else if $y_w(x) = 1$ and $y = 0$: $w \leftarrow w - \eta x$

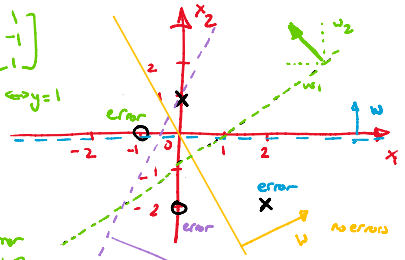
1 epoch = 1 pass over data

Simplifies to: $w \leftarrow w + \eta x (y - y(x))$

Theorem: If classes are linearly separable, will converge for any η.

Example: 1. Initialize $w = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

$$y_w(x) = \begin{cases} 1; & 1 - 1x_1 + 1x_2 \geq 0 \\ 0; & 1 - 1x_1 + 1x_2 < 0 \end{cases} \qquad x_2 \geq x_1 - 1 \Leftrightarrow y = 1$$



2. Epoch 1: (here, $\eta = 1$)
3. Updates:
   $w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \leftarrow w - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ (error) $\quad x_2 \geq 0 \Leftrightarrow y = 1$
   $w = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \leftarrow w + \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$ (error) $\quad 1 + 2x_1 - x_2 \geq 0$, $x_2 \leq 2x_1 + 1 \Leftrightarrow y = 1$
2. Epoch 2:
3. Updates: $w = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \leftarrow w - \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$ (error) $\quad 0 + 2x_1 + x_2 \geq 0$, $x_2 \geq -2x_1 \Leftrightarrow y = 1$

No more errors, terminate & return w.