CS 3/5780
Perceptron

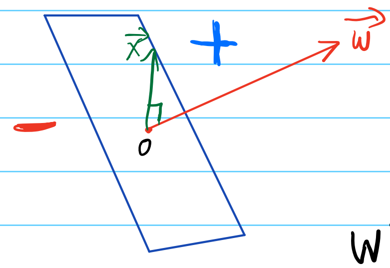
Rosenblatt '57

Assumption: the classes are "linearly separable"

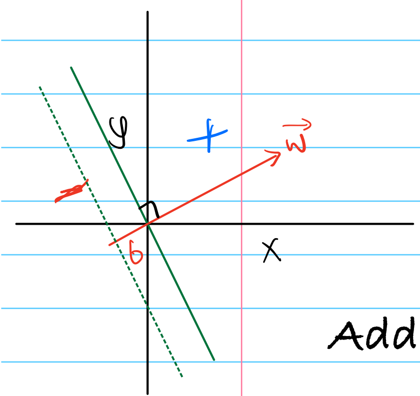
(positive and negative examples can be separated by a plane)

Binary Labels : $Y = \{+1, -1\}$

Inner product: $w, x \in \mathbb{R}^d$ $\quad w^T x = \sum_{i=1}^{d} w[i] \cdot x[i]$



$w^T x = 0$



$W^T X = \|W\| \, \|X\| \cos_{w,x}$

Add in bias: $y = w^T x + b$

$$h_{\vec{w},b}(\vec{x}) = \operatorname{sign}(\underbrace{\vec{w}}_{\text{weight vector}}{}^T \vec{x} + \underbrace{b}_{\text{bias}})$$

"Half space"

Absorbing bias by adding a dimension:

Quiz: By increasing dimension of features by 1 more, can you find a way to encode bias within just $\operatorname{sign}(w^T x)$?

$$\begin{bmatrix} \vec{w} \\ b \end{bmatrix}^T \begin{bmatrix} \vec{x} \\ 1 \end{bmatrix} = w^T x + b$$

After concatenating 1 to x's, use $h(x) = \text{sign}(w^T x)$

Using above we can always simply think about case without the bias term.

Perceptron Algorithm:

```
Initialize w = 0
While TRUE:
    M = 0                      # set no. of mistakes to 0
    For i = 1 to n :
        if y_i . w^T x_i <= 0      # is mistake
            w <- w + y_i x_i       # update
            M <- M + 1             # increment m
        endif
    END For
    if (M==0)
        BREAK;
    end if
END WHILE
```

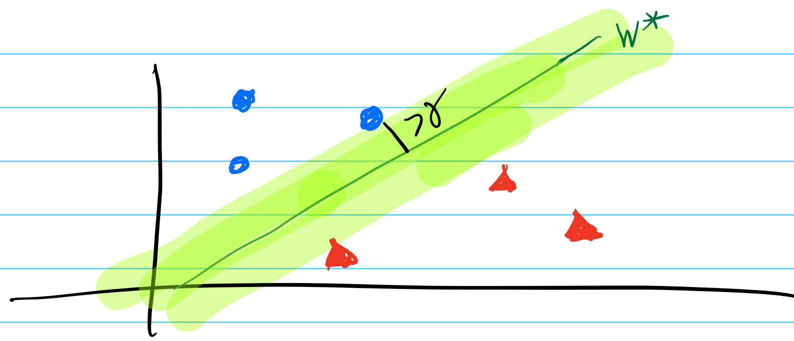
Update, $w \leftarrow w + y_i x_i$ improves on (x_i, y_i)

$$y_i (w + y_i x_i)^T x_i = y_i w^T x_i + x_i^T x_i$$

improvement

When will perceptron Algo. work?

when data is linearly Separable.



$\exists \vec{w}^*$ st

$\forall i \; y_i x_i^T w^* > 0$

$\|w^*\|_2 = 1$

Margin : $\gamma = \min_{i \in [n]} \; y_i \; x_i^T w^* > 0$ Assume: $\forall i, \; \|x_i\| \le 1$

How much time does it take to converge?

Thm: perceptron Algo. makes at most $1/\gamma^2$ updates.

Intuition: After each mistake we update as

$$w \leftarrow w + y_i x_i$$

1. After each update, $\|w\|^2$ increases by at most 1. Why?

Hence M updates implies $\|w\|^2 \le M$ or $\|w\| \le \sqrt{M}$

2. After each update, $w^T w^*$ increases by at least γ

Hence $w^T w^* \ge M\gamma$

Hence $M\gamma \le w^T w^* \le \|w\| \le \sqrt{M}$ and so $M \le \frac{1}{\gamma^2}$