Lec 4.1 PLA intuition

1. Recap:

- Decision Rule: $y = \text{sign}(\underline{w}^T \underline{x})$
- Augemented: $\underline{w} = (w_0, w_1, w_2 \dots w_d)$

  $\underline{x} = (x_0 = 1, x_1, x_2 \cdots w_d)$
- Preception Learning Algo:

  updated step: pick mis-classified example $(\underline{x}_n, y_n)$,

  $\underline{w} \leftarrow \underline{w} + y_n \underline{x}_n$

2. PLA intuition:

y_n	$\underline{w}^T \underline{x}_n$	correctly classfied?	$y_n \cdot \underline{w}^T \underline{x}_n$
+1	+	✓	+
+1	−	X	−
−1	+	X	−
−1	−	✓	+
0	0	x	0

- If $(\underline{x}_n, y_n)$ is correctly classfied. $y_n \underline{w}^T \underline{x}_n > 0$

  " " " misclassfied. $y_n \underline{w}^T \underline{x}_n \leq 0$

3. updating rule:

- Suppose $(\underline{x}_n, y_n)$ is mis-classified.

$$\underline{w}_{new} = \underline{w} + y_n \underline{x}_n$$

$$\therefore y_n \underline{w}_{new}^T \underline{x}_n = y_n (\underline{w} + y_n \underline{x}_n)^T \underline{x}_n$$

$$= y_n \underline{w}^T \underline{x}_n + \underbrace{y_n^2}_{+1} \underbrace{\|\underline{x}_n\|^2}_{>0}$$

$$> y_n \underline{w}^T \underline{x}_n$$

($\because$ Augmented $\underline{x}_n$ always has $x_{n0} = 1$)

↳. So this is a strict improvement for $\underline{x}_n$!

But could also cause new mis-classification for other data points.

↳ Thm: (Rosenblatt, 1957)

Given a linearly seperable dataset, PLA terminates in a finite No. of steps yielding $E_{in}(\underline{w}) = 0$

(prf: problem set 1.3)

↳. Remarks: The output of PLA is not unique

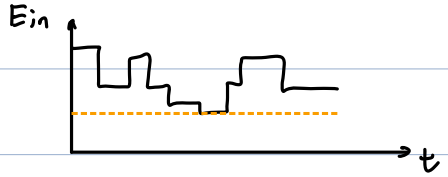
It depends on the initial line you choose and the order of classifed points.

Lec 4.2 Pocket Algorithm

1. Pocket Algo:

↳. Extends PLA for datasets that are not linearly seperable

↳ keep the "best" weight vector up to iteration t in the pocket.



↳. Algo step:

0'. pick time horizon T

1'. Set pocketed weight vector $\hat{\underline{w}}$ to $\underline{w}(0)$ of PLA.

2'. For $t = 0, 1, 2 \cdots, T-1$, do:

3'. Run PLA for one update to obtain $\underline{w}(t+1)$

4'. Evaluation $E_{in}(\underline{w}(t+1))$

5'. If $E_{in}(\underline{w}(t+1)) < E_{in}(\hat{\underline{w}})$

set $\hat{\underline{w}} = \underline{w}(t+1)$

6'. Return $\hat{\underline{w}}$

// No gartunee to find the optimal sol. This only heuristic algo. But perform well in practice.

4. Note:

Multiary classification can be implemented by a sequence of binary classfication

e.g. 10 digit example: