Perceptron criterion: $\min E_p(\mathbf{w}) = -\sum_{n \in \mathcal{M}} \mathbf{w}^T \mathbf{x}_n t_n - \sum_{n \in \mathcal{M}} \underbrace{\left(-\mathbf{w}^T \mathbf{x}_n t_n\right)}_{E_p(\mathbf{w}, \mathbf{x}_n)}$

where $\mathcal{M} = \{ 1 \leq n \leq N \mid \mathbf{w}^T \mathbf{x}_n t_n < 0 \}$ $\rightarrow$ misclassified.

Let's use SGD (1 example at a time).

minimize $E_p(\mathbf{w}, \mathbf{x}_n) = -\mathbf{w}^T \mathbf{x}_n t_n$

SGD update: $\mathbf{w} = \mathbf{w} - \eta \dfrac{\partial E_p}{\partial \mathbf{w}}$

$= \mathbf{w} - \eta \left(-\mathbf{x}_n t_n\right) = \mathbf{w} + \eta \mathbf{x}_n t_n$

- The value of η is immaterial (it does not change the label predicted by the model $\mathbf{w}^T\mathbf{x}$ that is learned with η).
- We can set $\eta = 1 \Rightarrow$ SGD update is $\boxed{\mathbf{w} = \mathbf{w} + \mathbf{x}_n t_n \text{ for any } \mathbf{x}_n \text{ that is } \mathcal{M}}$

1. $w \Leftarrow 0$

2. for $m = 1$ to N

3. if $w^T x_m t_m \leq 0$

4. $w \Leftarrow w + t_m x_m$

proof by MI that $\exists \alpha_n$ s.t. $w = \sum_{n=1}^{N} \alpha_n t_n x_n$, for any iteration of the for loop.

Base case: iteration not started yet (0), just evaluated line 1.

$$w = 0 = \sum_{n=1}^{N} \alpha_n t_n x_n \text{ where } \alpha_n = 0, \text{ for all } n = 1 \dots N.$$

inductive step: Assume $w = \sum_{n=1}^{N} \alpha_n t_n x_n$ when $m = k \geq 0$

We want to prove that $(\exists)$ α'_n s.t. $w = \sum_{n=1}^{N} \alpha'_n t_n x_n$ when $m = k+1$.

Case 1: the model w does not make a mistake on x_k.

then $w = \sum_{n=1}^{N} \alpha'_n t_n x_n$, where $\alpha'_n = \alpha_n$

The current perceptron model w makes a mistake on example x_n.

$\Rightarrow$ $\underline{w^T x_n t_n} < 0$

$\Rightarrow$ w is updated to $\bar{w} = w + t_n x_n$

Then $\underline{\bar{w}^T x_n t_n} = (w + t_n x_n)^T x_n t_n = \underline{w}^T x_n t_n + t_n x_n^T x_n t_n$

$$= w^T x_n t_n + t_n^2 \|x_n\|^2$$

$$= w^T x_n t_n + \underbrace{\|x_n\|^2}_{\geq 0}$$

$$\geq w^T x_n t_n$$

Case 2: w makes a mistake on x_k

$$\Rightarrow W = w + x_k t_k = \sum_{n=1}^{N} \underline{\alpha_n} t_n x_n + x_k t_k$$

$$= \sum_{n=1}^{N} \alpha'_n t_n x_n,$$

where $\alpha'_n = \begin{cases} \alpha_n + 1, & \text{if } n = k \\ \alpha_n, & \text{if } n \neq k \end{cases}$

(q.e.d.).

$$w = \underbrace{\sum_{n=1}^{N} \alpha_n t_n x_n}, \qquad \text{start w/ } \alpha_n = 0 \Longleftrightarrow w = 0$$

1. Set $\vec{\alpha} \leftarrow 0$ $\qquad \alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$

2. for $m = 1$ to N

3. if $w^T x_m t_m \le 0$: $\Longleftrightarrow$ if $\sum_{n=1}^{N} \alpha_n t_n \underbrace{\overbrace{x_n^T x_m}^{K(x_n, x_m)}} t_m \le 0$ $\qquad (x_n^T x_m + 1)^2$

4. $w \leftarrow w + x_m t_m \Longleftrightarrow \alpha_m \leftarrow \alpha_m + 1$

5. return $\vec{w}$ $\qquad$ return $\vec{\alpha}$