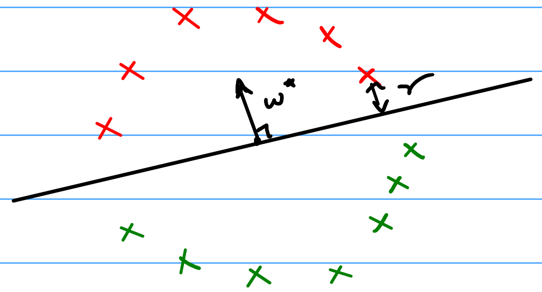
Correction TD5

Exercise 9 : Perceptron with margin



$$\gamma = \sup_{\substack{w \in \mathbb{R}^d \\ \|w\|=1}} \min_{1 \leq i \leq n} \frac{y_i \langle w, x_i \rangle}{\|x_i\|}$$

1. By definition, $\exists (w_t)$ s.t. $\forall t$, $\|w_t\| = 1$, $\displaystyle \min_{1 \leq i \leq n} \frac{y_i \langle w_t, x_i \rangle}{\|x_i\|} \xrightarrow[n \to +\infty]{} \gamma$

$\{w : \|w\| = 1\}$ is compact. Indeed, it is bounded by definition and is equal to $\|\cdot\|^{-1}(\{1\})$ so it is a closed set.
Since we work in finite dimension, this proves the compactness

Hence, $\exists \varphi : \mathbb{N} \to \mathbb{N}$ strictly increasing such that $w_{\varphi(t)} \to w^*$ and $\|w^*\| = 1$

Furthermore, $\displaystyle w \mapsto \min_{1 \leq i \leq n} \frac{y_i \langle w_n, x_i \rangle}{\|x_i\|}$ is continuous,

So, $\displaystyle \gamma = \min_{1 \leq i \leq n} \frac{y_i \langle w^*, x_i \rangle}{\|x_i\|}$

So, $\forall 1 \leq i \leq n$, $\displaystyle \frac{y_i \langle w^*, x_i \rangle}{\|x_i\|} \geq \gamma$

2 Perceptron with margin γ :

Input : Margin γ

Data : training set $(x_1, y_1), \dots, (x_n, y_n)$

$w_0 \leftarrow (0, \dots, 0)$

$t \geqslant 0$

while $\exists\, i_t : y_{i_t} \langle w_t, x_{i_t} \rangle \leqslant \frac{\gamma}{2} \|x_{i_t}\| \|w_t\|$ do

$$w_{t+1} = w_t + y_{i_t} \frac{x_{i_t}}{\|x_{i_t}\|}$$

$t \leftarrow t+1$

return w_t

Let's prove that $\forall t,\ \langle w^*, w_t \rangle \geqslant \gamma t$

• $t=0$ $\langle w^*, w_0 \rangle = 0 = 0\gamma$

• Let's supose that $\langle w^*, w_{t-1} \rangle \geqslant \gamma(t-1)$ for a given $t \geqslant 1$ and w_t passed the test.

$$w_t = w_{t-1} + y_{i_{t-1}} \frac{x_{i_{t-1}}}{\|x_{i_{t-1}}\|}$$

So, $\langle w^*, w_t \rangle = \langle w^*, w_{t-1} \rangle + y_{i_{t-1}} \langle w^*, x_{i_{t-1}} \rangle / \|x_{i_{t-1}}\|$

$$\geqslant \underbrace{\gamma(t-1)}_{\text{rec}} + \underbrace{\gamma}_{w^*}$$

$$\geqslant \gamma t$$

Which proves that $\forall t$ that passes the test, $\langle w^*, w_t \rangle \geqslant \gamma t$

3. $\|w_{t+1}\|^2 = \|w_t + y_{i_t}\frac{x_{i_t}}{\|x_{i_t}\|}\|^2$

$$= \|w_t\|^2 + \underbrace{2 y_{i_t} \langle w_t, x_{i_t}\rangle / \|x_{i_t}\|}_{\leq \|w_t\| \frac{\gamma}{2} \text{ by test}} + y_{i_t}^2 \frac{\|x_{i_t}\|^2}{\|x_{i_t}\|^2}$$

$$\leq \|w_t\|^2 + \gamma \|w_t\| + 1$$

So, $\boxed{\|w_{t+1}\|^2 \leq \|w_t\|^2 + \gamma\|w_t\| + 1}$

4. So, $\|w_{t+1}\|^2 \leq \left(\|w_t\| + \frac{3\gamma}{4}\right)^2 + 1 - \frac{\gamma}{2}\|w_t\|$

And if $\|w_t\| \geq 2/\gamma$, $\|w_{t+1}\|^2 \leq \left(\|w_t\| + \frac{3\gamma}{4}\right)^2$

5. Let's prove it by recurrence:

- $t=0$ $\|w_0\| = 0 \leq 1 + \frac{2}{\gamma} + \frac{3\gamma t}{4}$

- Let's suppose the result true for a given t.

  * if $\|w_t\| < 2/\gamma$,

$$\|w_{t+1}\| = \|w_t + y_{i_t}\frac{x_{i_t}}{\|x_{i_t}\|}\| \leq \|w_t\| + |y_{i_t}| \frac{\|x_{i_t}\|}{\|x_{i_t}\|}$$

$$\leq 2/\gamma + 1 \leq 1 + \frac{2}{\gamma} + \frac{3\gamma(t+1)}{4}$$

- if $\|w_t\| \geq 2/\gamma$

$$\|w_{t+1}\| \leq \|w_t\| + \frac{3\gamma}{4}$$

$$\leq 1 + \frac{2}{\gamma} + \frac{3\gamma t}{4} + \frac{3\gamma}{4} \quad \text{by rec}$$

$$\leq 1 + \frac{2}{\gamma} + \frac{3\gamma(t+1)}{4}$$

So $\forall t, \|w_t\| \leq 1 + \frac{2}{\gamma} + \frac{3\gamma t}{4}$

6. So, As long as we stay in the loop, we have,

$$\begin{cases} \gamma t \leq \langle w^*, w_t \rangle \underset{\text{C.S.}}{\leq} \|w^*\| \|w_t\| = \|w_t\| \\ \|w_t\| \leq 1 + \frac{2}{\gamma} + \frac{3\gamma t}{4} \end{cases}$$

And so we are sure to have a break before

$$\gamma t \geq 1 + \frac{2}{\gamma} + \frac{3\gamma t}{4}$$

Furthermore, $0 < \gamma \leq 1$ so $t \geq 12/\gamma^2$ is a sufficient condition.

Conclusion: $\boxed{\text{The algorithm achieves margin at least } \gamma/2 \text{ in at most } 12/\gamma^2 \text{ iterations}}$

7. Condition becomes

$$\boxed{\text{while } \exists i_t : y_{i_t} \langle w_t, x_{i_t} \rangle \leq (1-\eta)\gamma \|x_{i_t}\| \|w_t\|}$$

- $\langle w^*, w_t \rangle \geq \gamma t$ Still true
- $\|w_{t+1}\|^2 \leq \|w_t\|^2 + 2(1-\eta)\gamma\|w_t\| + 1$
- if $\|w_t\| \geq \frac{1}{\gamma\eta}$, $\|w_{t+1}\|^2 \leq \left(\|w_t\| + \frac{2-\eta}{2}\gamma\right)^2$
- $\|w_t\| \leq 1 + \frac{1}{\gamma\eta} + \frac{(2-\eta)}{2}\gamma t$

So we know that it stops when $\gamma t \geq 1 + \frac{1}{\gamma\eta} + \frac{(2-\eta)}{2}\gamma t$

So $t \geq \frac{1}{\gamma^2}\left[\frac{2}{\eta}\left(1 + \frac{1}{\eta}\right)\right]$ is enough.

The algorithm produces a margin at least $(1-\eta)\gamma$ in at most $K(\eta)/\gamma^2$ operations with

$$K(\eta) = \frac{2}{\eta}\left(1 + \frac{1}{\eta}\right)$$