The Perceptron Algorithm

(STOP in Finite Steps)

Theorem (Novikoff)

Let S be a non-trivial training set, and let $R = \max\limits_{1 \leqslant i \leqslant \ell} ||x^i||_2$.

Suppose that there exists a vector $w_{opt} \in R^n, ||w_{opt}|| = 1$ and $y_i(\langle w_{opt} \cdot x^i \rangle + b_{opt}) \geqslant \gamma, \forall \ 1 \leqslant i \leqslant \ell$. Then the number of mistakes made by the on-line perceptron algorithm on S is at most $(\frac{2R}{\gamma})^2$.

Proof of Finite Termination

Proof: Let

$$\overline{x}^i = \begin{bmatrix} x^i \\ R \end{bmatrix}, \overline{w} = \begin{bmatrix} w \\ \frac{b}{R} \end{bmatrix}.$$

The algorithm starts with an augmented weight vector $\overline{w}^0 = \mathbf{0}$ and updates it at each mistake.

Let $\overline{w}^{t-1}$ be the augmented weight vector prior to the t th mistake. The t th update is performed when

$$y_i \langle \overline{w}^{t-1}, \overline{x}^i \rangle = y_i (\langle w^{t-1}, x^i \rangle + b_{t-1}) \leqslant 0$$

where $(x^i, y_i) \in S$ is the point incorrectly classified by $\overline{w}^{t-1}$.

Update Rule of Perceotron

$$w^t \leftarrow w^{t-1} + \eta y_i x^i \;\; ; b_t \leftarrow b_t + \eta y_i R^2$$

$$\overline{w}^t = \begin{bmatrix} w^t \\ \frac{b_t}{R} \end{bmatrix} = \begin{bmatrix} w^{t-1} \\ \frac{b_{t-1}}{R} \end{bmatrix} + \eta y_i \begin{bmatrix} x^i \\ R \end{bmatrix} = \overline{w}^{t-1} + \eta y_i \overline{x}^i.$$

$$\langle \overline{w}^t, \overline{w}_{opt} \rangle = \langle \overline{w}^{t-1}, \overline{w}_{opt} \rangle + \eta y_i \langle \overline{x}^i, \overline{w}_{opt} \rangle$$

$$\geqslant \langle \overline{w}^{t-1}, \overline{w}_{opt} \rangle + \eta\gamma \geqslant \langle \overline{w}^{t-2}, \overline{w}_{opt} \rangle + 2\eta\gamma \ldots \geqslant t\eta\gamma$$

$$(NOTE : y_i(\langle w_{opt} \cdot x^i \rangle + b_{opt}) \geqslant \gamma, \forall \;\; 1 \leqslant i \leqslant \ell)$$

Similarly, $||\overline{w}^t||_2^2 = ||\overline{w}^{t-1}||_2^2 + 2\eta y_i \langle \overline{w}^{t-1}, \overline{x}^i \rangle + \eta^2 ||\overline{x}^i||_2^2$

$$\leqslant ||\overline{w}^{t-1}||_2^2 + \eta^2 ||\overline{x}^i||_2^2 = ||\overline{w}^{t-1}||_2^2 + \eta^2 (||x^i||_2^2 + R^2)$$

$$\leqslant ||\overline{w}^{t-2}||_2^2 + 2\eta^2 R^2 \leqslant \cdots \leqslant 2t\eta^2 R^2$$

Update Rule of Perceotron

$$w^t \leftarrow w^{t-1} + \eta y_i x^i \;\; ; b_t \leftarrow b_t + \eta y_i R^2$$

$$\langle \overline{w}^t, \overline{w}_{opt} \rangle \geqslant t\eta\gamma \;\; and \;\; ||\overline{w}^t||_2^2 \leqslant 2t\eta^2 R^2$$

$$||\overline{w}_{opt}||_2 \sqrt{2t}\, \eta R \geqslant ||\overline{w}_{opt}||_2 ||\overline{w}^t||_2 \geqslant \langle \overline{w}^t, \overline{w}_{opt} \rangle \geqslant t\eta\gamma$$

$$\Rightarrow \;\; t \leqslant 2\left(\frac{R}{\gamma}\right)^2 ||\overline{w}_{opt}||_2^2 \leqslant \left(\frac{2R}{\gamma}\right)^2$$

$$Note: \;\; b_{opt} \leqslant R, \;\; ||\overline{w}_{opt}||_2^2 \leqslant ||w_{opt}||_2^2 + 1 = 2$$