

Lectures

Ilia Kamyshev

January 8, 2021

Contents

1	Rosenblatt's perceptron	1
1.1	Introduction	1

1 Rosenblatt's perceptron

1.1 Introduction

Recall: for an input $x \in \mathbb{R}^n$, the parametrized function describing the mapping computed by a perceptron is:

$$F(\omega, x) = F((\omega_1, \omega_2, \dots, \omega_n, \theta), (x_1, x_2, \dots, x_n, -1)) = \text{sgn}\left(\sum_{i=1}^n \omega_i x_i - \theta\right)$$

Or in the vector form:

$$F(\omega, \mathbf{x}) = \text{sgn}(\omega^T \mathbf{x})$$

This perceptron can only classify patterns which are *linearly separable*.

Theorem 1. Suppose $C_1 \cup C_2 = C$ are linearly separable classes over the training set $z \in Z^T$ with the assumption $z_t = \{\mathbf{x}_t, y_t \in C_1\}$, and perceptron's response r_t with mistake $e_t = r_t - y_t < 0$ can be corrected by applying the learning rule to its current state $\omega \in \Omega^T$:

$$\omega_t = \omega_{t-1} + e_t \mathbf{x}_t$$

Then perceptron's error correction algorithm converges in k number of steps with following assumptions: training input is bounded by Euclidean norm $\|\mathbf{x}_t\| \leq R$ and $e_t \omega_*^T \mathbf{x}_t \geq \gamma$ for $t = 1..T$, where $\gamma > 0$. Initial state $\omega_0 = 0$. Note, that γ uses to be sure that some example is classified correctly.

Proof: Multiplying both sides of learning rule equation by some optimal ω_*^T we will have:

$$\omega_*^T \omega_k = \omega_*^T \omega_{k-1} + e_k \omega_*^T x_k \geq \omega_*^T \omega_{k-1} + \gamma$$

Now we can expand equation above for k steps and keep in mind $\omega_0 = 0$ we will get:

$$\begin{aligned} \omega_*^T \omega_k &\geq \omega_*^T (\omega_{k-2} + e_{k-1} x_{k-1}) + \gamma \geq \omega_*^T (\omega_{k-3} + e_{k-2} x_{k-2}) + 2\gamma \geq \dots \\ &\dots \geq \omega_*^T (\omega_0 + e_1 x_1) + (k-1)\gamma \geq k\gamma \end{aligned}$$

Let's do one important step which results will be substituted to the final inequality. Suppose we have following Euclidean norm $\|\omega_k\|$ and, as it's known, for squared $L2$ norm the equality holds:

$$\|\omega_k\|^2 = \|\omega_{k-1} + e_k x_k\|^2 = \|\omega_{k-1}\|^2 + e_k^2 \|x_k\|^2 + 2e_k \omega_{k-1}^T x_k$$

Since $e_k = r_k - y_k = -2$ ($r_k = -1$ while the target $y_k = 1$):

$$\|\omega_{k-1}\|^2 + 2\|x_k\|^2 - 4\omega_{k-1}^T x_k$$

There is no doubt that:

$$\|\omega_{k-1}\|^2 + 2\|x_k\|^2 - 4\omega_{k-1}^T x_k \leq \|\omega_{k-1}\|^2 + 2\|x_k\|^2$$

Continuing for k steps we will have:

$$\|\omega_{k-1}\|^2 + 2\|x_k\|^2 - 4\omega_{k-1}^T x_k \leq 2 \sum_{j=1}^k \|x_j\|^2$$

Since $\|x_t\| \leq R$:

$$\|\omega_{k-1}\|^2 + 2\|x_k\|^2 - 4\omega_{k-1}^T x_k \leq 2kR^2$$

From *Cauchy-Schwarz inequality*:

$$|\omega_*^T \omega_k| \leq \|\omega_*^T\| \cdot \|\omega_k\|$$

$$k\gamma \leq \|\omega_*^T\| \cdot \|\omega_k\|$$

$$\frac{k\gamma}{\|\omega_*^T\|} \leq \|\omega_k\|$$

Substituting results obtained earlier:

$$\frac{k^2\gamma^2}{\|\omega_*^T\|^2} \leq \|\omega_k\|^2 = \|\omega_{k-1}\|^2 + 2\|x_k\|^2 - 4\omega_{k-1}^T x_k \leq 2kR^2$$

Finally:

$$\begin{aligned} \frac{k^2\gamma^2}{\|\omega_*^T\|^2} &\leq 2kR^2 \\ \frac{k\gamma^2}{\|\omega_*^T\|^2} &\leq 2R^2 \\ k &\leq \frac{2R^2}{\gamma^2} \|\omega_*^T\|^2 \blacksquare \end{aligned}$$