Lectures

Ilia Kamyshev

January 8, 2021

Contents

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:

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

Or in the vector form:

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

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

Theorem 1. *Suppose* $C_1 \bigcup C_2 = C$ *are linearly separable classes over the training set* $z \in Z^T$ *with the assumption* $z_t = \{\boldsymbol{x_t}, y_t \in C_1\}$*, 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* $\boldsymbol{\omega} \in \boldsymbol{\Omega}^T$:

$$\boldsymbol{\omega_t} = \boldsymbol{\omega_{t-1}} + e_t \boldsymbol{x_t}$$

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

Proof: Multiplying both sides of learning rule equation by some optimal $\boldsymbol{\omega_*^T}$ we will have:

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

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

$$\boldsymbol{\omega_*^T \omega_k} \geq \boldsymbol{\omega_*^T}(\boldsymbol{\omega_{k-2}} + e_{k-1}\boldsymbol{x_{k-1}}) + \gamma \geq \boldsymbol{\omega_*^T}(\boldsymbol{\omega_{k-3}} + e_{k-2}\boldsymbol{x_{k-2}}) + 2\gamma \geq \dots$$

$$\dots \geq \boldsymbol{\omega_*^T}(\boldsymbol{\omega_0} + e_1 x_1) + (k-1)\gamma \geq k\gamma$$

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

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

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

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

There is no doubt that:

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

Continuing for k steps we will have:

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

Since $\|\boldsymbol{x_t}\| \leq R$:

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

From *Cauchy-Schwarz inequality*:

$$\left|\boldsymbol{\omega_*^T \omega_k}\right| \leq \left\|\boldsymbol{\omega_*^T}\right\| \cdot \|\boldsymbol{\omega_k}\|$$

$$k\gamma \leq \left\|\boldsymbol{\omega_*^T}\right\| \cdot \|\boldsymbol{\omega_k}\|$$

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

Substitying results obtained earlier:

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

Finally:

$$\frac{k^2\gamma^2}{\left\|\boldsymbol{\omega_*^T}\right\|^2} \leq 2kR^2$$

$$\frac{k\gamma^2}{\left\|\boldsymbol{\omega_*^T}\right\|^2} \leq 2R^2$$

$$k \leq \frac{2R^2}{\gamma^2}\left\|\boldsymbol{\omega_*^T}\right\|^2 \blacksquare$$