The perceptron and large margin classifiers

February 4, 2021

1 The perceptron and large margin classifiers

Online learning: The algorithm has to make predictions continuously even when it is learning.
Perceptron algorithm has parameters $\theta \in \mathcal{R}^{n+1}$ (i.e. that means, it has an offset as well, i.e. distance of the decision boundary from origin.). Its hypothesis is,

$$h_\theta(x) = g\left(\theta^T x\right)$$

where,

$$g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

. Parameters update: if $h_\theta(x) = y$, then it makes no change to the parameters, otherwise it performs the following update,

$$\theta := \theta + yx$$

1.1 Bound on errors of perceptron algorithm

Let a sequence of examples $\left(x^{(1)}, y^{(1)}\right), \left(x^{(2)}, y^{(2)}\right), \ldots \left(x^{(m)}, y^{(m)}\right)$ is given.
Suppose that $\left\|x^{(i)}\right\| \leq D$ for all i.
There exist a unit vector $u\left(\|u\|_2 = 1\right)$ such that $y^{(i)} \cdot \left(u^T x^{(i)}\right) \geq \gamma$ for all examples in sequence. i.e. Margin γ which separates all the examples perfectly.
Then, total number of mistakes that the perceptron algorithm makes on this sequence is at most $(D/\gamma)^2$ **Proof.** Whenever there is an error:

$$\left(x^{(i)}\right)^T \theta^{(k)} y^{(i)} \leq 0$$

Then we have parameter update as follow:

$$\theta^{(k+1)} = \theta^{(k)} + y^{(i)} x^{(i)}$$

Since u gives a perfect decision boundary, with the minimum margin γ.

$$\begin{aligned} \left(\theta^{(k+1)}\right)^T u &= \left(\theta^{(k)}\right)^T u + y^{(i)} \left(x^{(i)}\right)^T u \\ &\geq \left(\theta^{(k)}\right)^T u + \gamma \end{aligned}$$

By a straightforward induction, we can imply,

$$\left(\theta^{(k+1)}\right)^T u \geq k\gamma$$

Also we have,

$$\begin{aligned} \left\|\theta^{(k+1)}\right\|^2 &= \left\|\theta^{(k)} + y^{(i)} x^{(i)}\right\|^2 \\ &= \left\|\theta^{(k)}\right\|^2 + \left\|x^{(i)}\right\|^2 + 2y^{(i)} \left(x^{(i)}\right)^T \theta^{(i)} \\ &\leq \left\|\theta^{(k)}\right\|^2 + \left\|x^{(i)}\right\|^2 \\ &\leq \left\|\theta^{(k)}\right\|^2 + D^2 \end{aligned}$$

Note, in the second step listed above, $y^{(i)}\left(x^{(i)}\right)^T\theta^{(i)} < 0$ because of the error. With inductive argument, we can imply,

$$\left\|\theta^{(k+1)}\right\|^2 \leq kD^2$$

. Hence, we can put together as,

$$\begin{aligned}\sqrt{k}D &\geq \left\|\theta^{(k+1)}\right\| \\ &\geq \left(\theta^{(k+1)}\right)^T u \\ &\geq k\gamma\end{aligned}$$

. Hence max mistakes a perceptron can make is, $k \leq (D/\gamma)^2$