

Perceptron

Friday, April 12, 2024

8:09 PM

Assume we have a classification problem.
Given $x \in \mathbb{R}^D$, we want $y \in \{0, 1\}$.
We solve it the following way

$$x \mapsto w^T x + b \mapsto \sigma(w^T x + b) =: \hat{y}$$

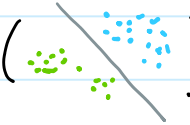
(weights from training) activation function

Update Rule of weights

$$w \in \mathbb{R}^D.$$

$$w_{\text{new}} := w_{\text{old}} + \underset{\substack{\text{LR} \\ \downarrow \\ \text{training samples } x \in \mathbb{R}^D}}{\alpha} \sum (y_i - \hat{y}_i) x$$

Intuitively, why does this solve?

If our data is linearly separable (, then this converges.

As σ we can take

