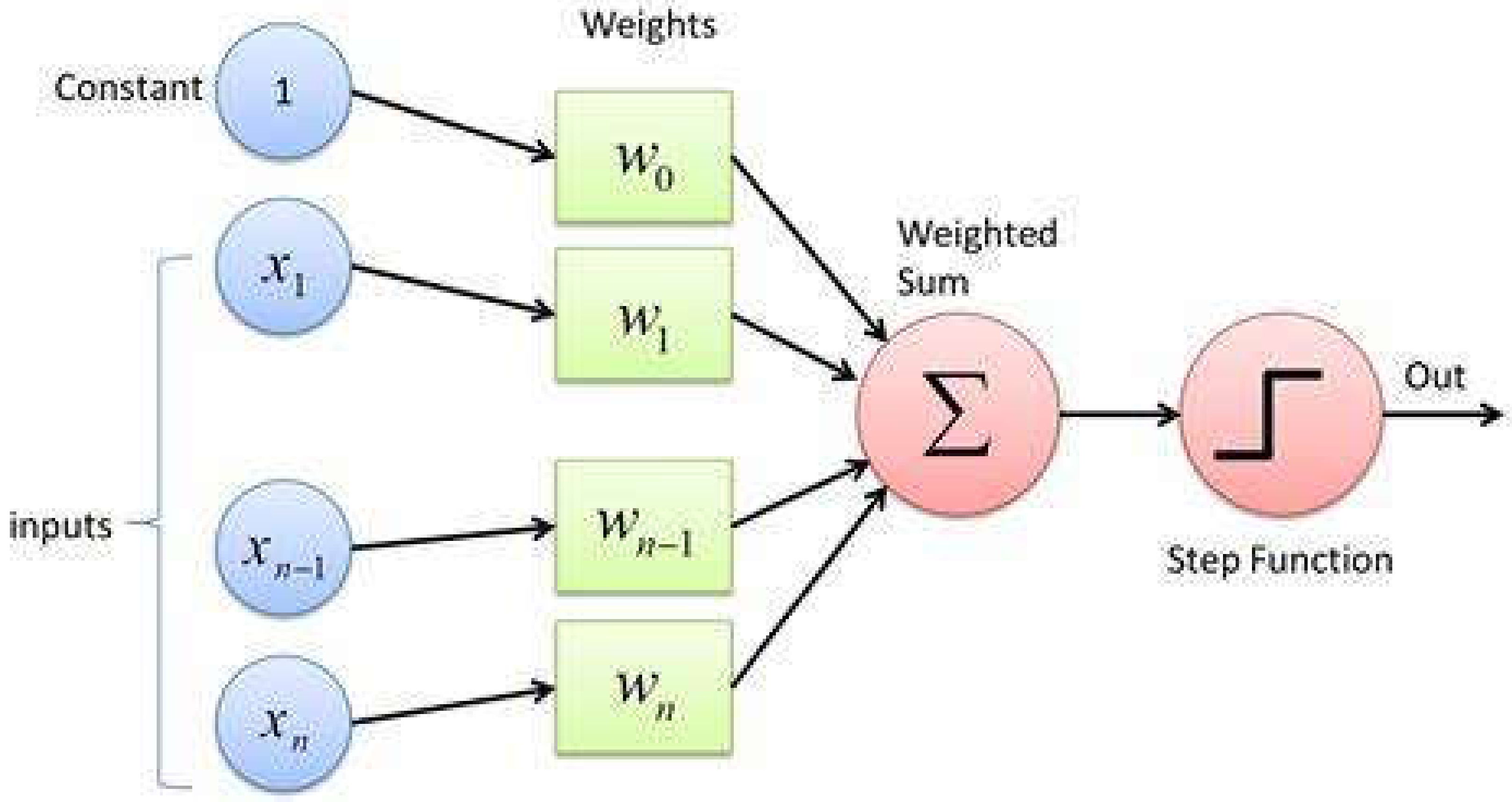
感知机

Peng Li

https://simplelp.github.io/

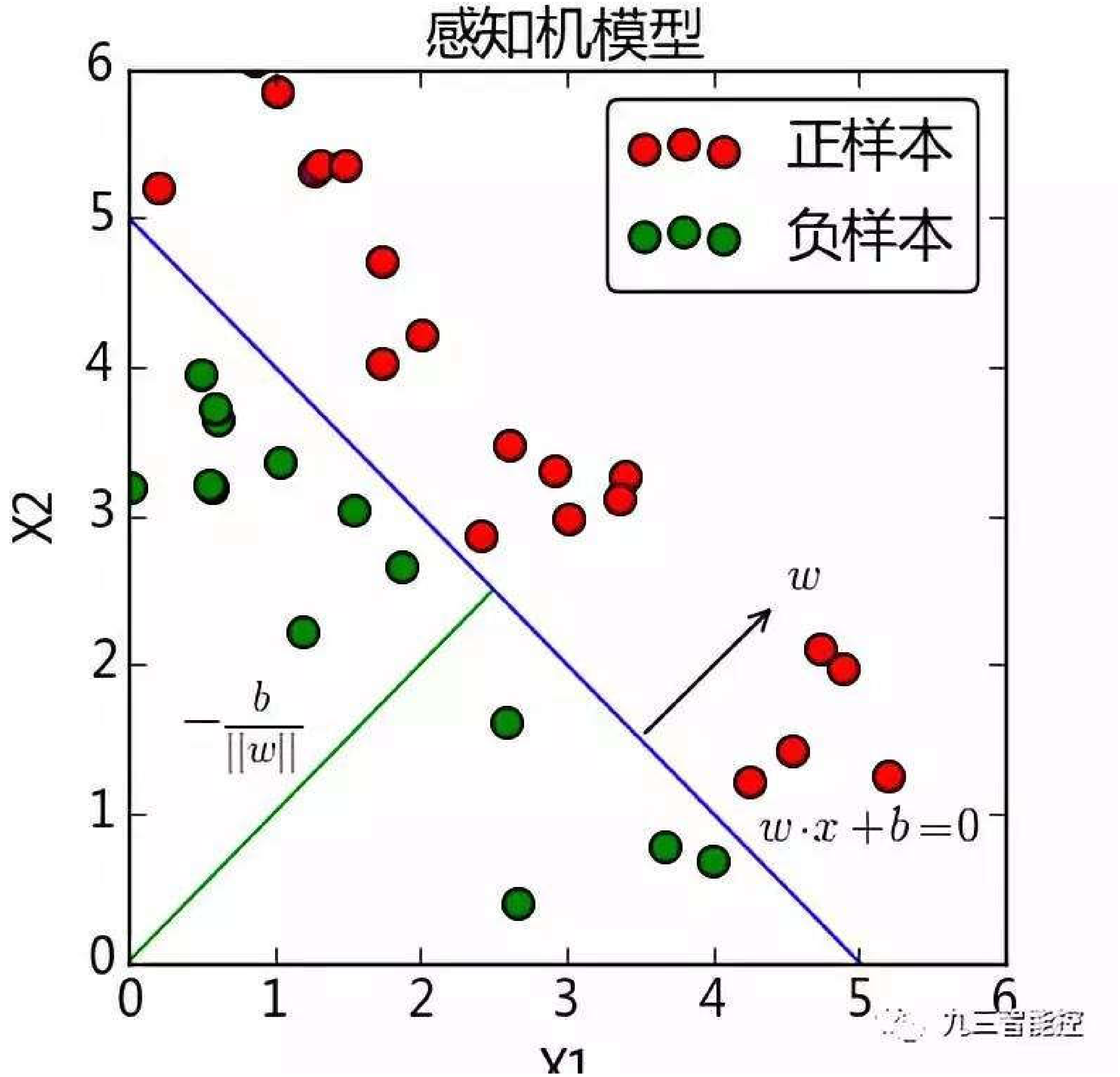
2019/06/03

感知机模型



Weights
Constant
1
w_0
x_1
w_1
Weighted Sum
Σ
Out
inputs
x_{n-1}
w_{n-1}
Step Function
x_n
w_n

感知机模型



感知机模型学习策略

- 对于线性可分的训练集中的误分类点：

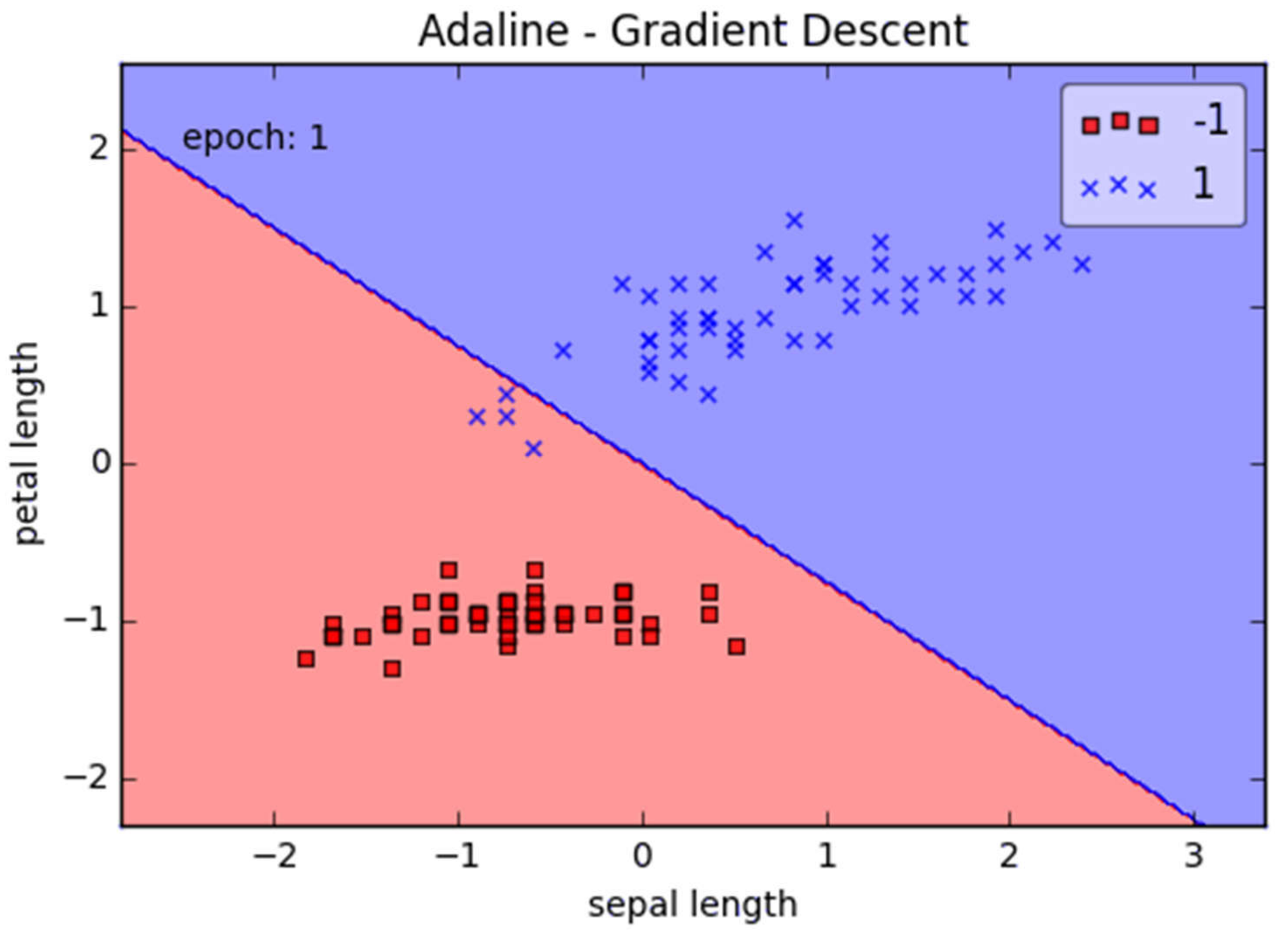
$$-\sum_{x_i \in M} \frac{y_i(w \cdot x_i + b)}{\|w\|}$$

感知机模型学习策略

- 感知机损失函数

$$L(w, b) = -\sum_{x_i \in M} y_i (w \cdot x_i + b)$$

感知机学习算法


Adaline - Gradient Descent
epoch: 1
-1
1
2
1
0
−1
−2
−2
−1
0
1
2
3
petal length
sepal length

感知机学习算法

算法 2.1（感知机学习算法的原始形式）

输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in\mathcal{X}=\mathbf{R}^n$，$y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$；学习率 $\eta\,(0<\eta\leqslant 1)$；

输出：w,b；感知机模型 $f(x)=\operatorname{sign}(w\cdot x+b)$.

（1）选取初值 w_0,b_0

（2）在训练集中选取数据 (x_i,y_i)

（3）如果 $y_i(w\cdot x_i+b)\leqslant 0$

$$w\leftarrow w+\eta y_i x_i$$

$$b\leftarrow b+\eta y_i$$

（4）转至（2），直至训练集中没有误分类点. ∎

感知机学习算法

定理 2.1（Novikoff） 设训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$ 是线性可分的，其中 $x_i\in\mathcal{X}=\mathbf{R}^n$，$y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$，则

（1）存在满足条件 $\|\hat{w}_{\mathrm{opt}}\|=1$ 的超平面 $\hat{w}_{\mathrm{opt}}\cdot\hat{x}=w_{\mathrm{opt}}\cdot x+b_{\mathrm{opt}}=0$ 将训练数据集完全正确分开；且存在 $\gamma>0$，对所有 $i=1,2,\cdots,N$

$$y_i(\hat{w}_{\mathrm{opt}}\cdot\hat{x}_i)=y_i(w_{\mathrm{opt}}\cdot x_i+b_{\mathrm{opt}})\geqslant\gamma \tag{2.8}$$

（2）令 $R=\max\limits_{1\leqslant i\leqslant N}\|\hat{x}_i\|$，则感知机算法 2.1 在训练数据集上的误分类次数 k 满足不等式

$$k\leqslant\left(\frac{R}{\gamma}\right)^2 \tag{2.9}$$

感知机学习算法

算法 2.2（感知机学习算法的对偶形式）

输入：线性可分的数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in\mathbf{R}^n$， $y_i\in\{-1,+1\}$， $i=1,2,\cdots,N$；学习率 η（$0<\eta\leqslant 1$）；

输出：α,b；感知机模型 $f(x)=\mathrm{sign}\left(\sum_{j=1}^{N}\alpha_j y_j x_j\cdot x+b\right)$.

其中 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_N)^{\mathrm{T}}$.

（1）$\alpha\leftarrow 0$， $b\leftarrow 0$

（2）在训练集中选取数据 (x_i,y_i)

（3）如果 $y_i\left(\sum_{j=1}^{N}\alpha_j y_j x_j\cdot x_i+b\right)\leqslant 0$

$$\alpha_i\leftarrow\alpha_i+\eta$$
$$b\leftarrow b+\eta y_i$$

（4）转至（2）直到没有误分类数据. ■

思考题

- 如何从加权、向量乘法的角度理解感知机模型?
- 一个点到超平面公式的证明?
- 为什么感知机学习可以用函数间隔?
- 感知机学习算法有没有公式解?
- 如何利用python实现感知机模型?
- 负梯度方向是损失函数降低最快的方向的证明?
- 证明感知机无法解决异或问题?
- 证明Novikoff定理?

Thanks !