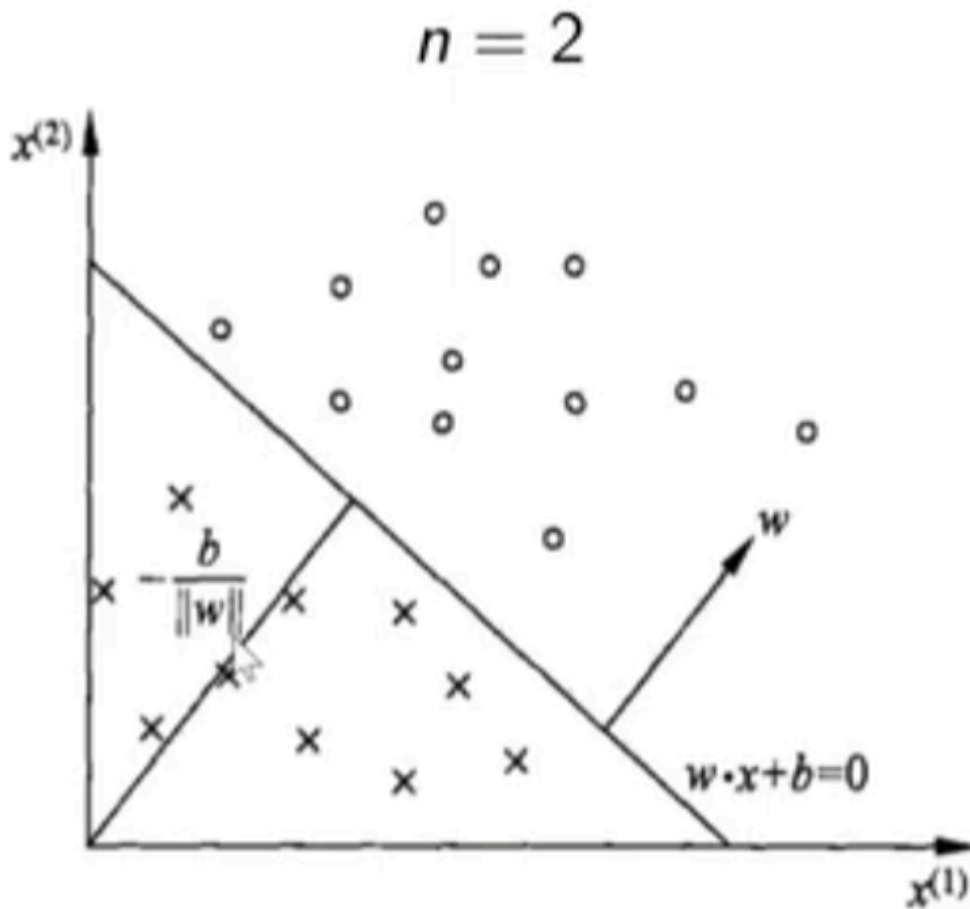


感知机是一类最简单的线性分类模型，与之前的logistic regression相似但是是弱化版。

感知机的原理是对一个可线性划分的空间找到其超平面，使得这个超平面可以将所有样本点划分为完全分开的两类。表达为：

对于输入 $x \in R^N$,找出 \hat{w} 与 \hat{b} ，使得平面被直线 $y = \hat{w}x + \hat{b}$ 划分为两部分，由线性规划得到这个输出值小于0的将被分为负类，而大于零的将被分为正类。

如图：



对于训练的损失函数，我们选择分类错误的样本点的距离：

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

之后依据梯度下降法即可训练找到合适的参数 \hat{w} 与 \hat{b} 。

原始形式：随机梯度下降法

Janneil博士 bilibili

损失函数：

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

梯度：

$$\nabla_w L(w, b) = - \sum_{x_i \in M} y_i x_i; \quad \nabla_b L(w, b) = - \sum_{x_i \in M} y_i$$

参数更新：

- 批量梯度下降法（Batch Gradient Descent）：每次迭代时使用所有误分类点来进行参数更新。

$$w \leftarrow w + \eta \sum_{x_i \in M} y_i x_i; \quad b \leftarrow b + \eta \sum_{x_i \in M} y_i$$

其中， $\eta(0 < \eta \leq 1)$ 代表步长。

- 随机梯度下降法（Stochastic Gradient Descent）：每次随机选取一个误分类点。

$$w \leftarrow w + \eta y_i x_i; \quad b \leftarrow b + \eta y_i$$

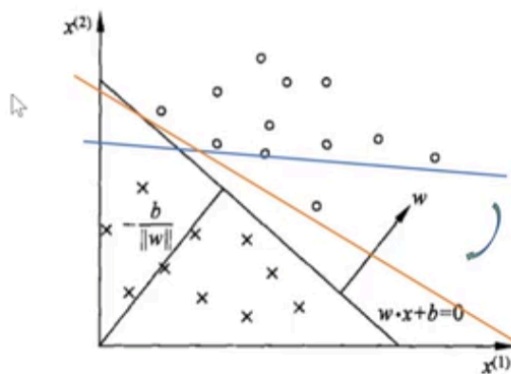
输入：训练集：

$$T = \{(x_1, y_1), (x_2, y_2) \cdots, (x_N, y_N)\}$$

其中， $x_i \in \mathcal{X} \subseteq \mathbf{R}^n$ ， $y \in \mathcal{Y} = \{+1, -1\}$ ；步长 $\eta (0 < \eta \leq 1)$

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

- (1) 选取初始值 w_0, b_0 ;
- (2) 于训练集中随机选取数据 (x_i, y_i) ;
- (3) 若 $y_i(w \cdot x_i + b) \leq 0$,
 $w \leftarrow w + \eta y_i x_i$; $b \leftarrow b + \eta y_i$
- (4) 转(2)，直到训练集中没有误分类点。



感知机模型的缺陷也很明显，就是功能过弱，只能对可线性划分的样本空间进行操作。之后的很多模型就是建立在感知机的基础上诞生的。