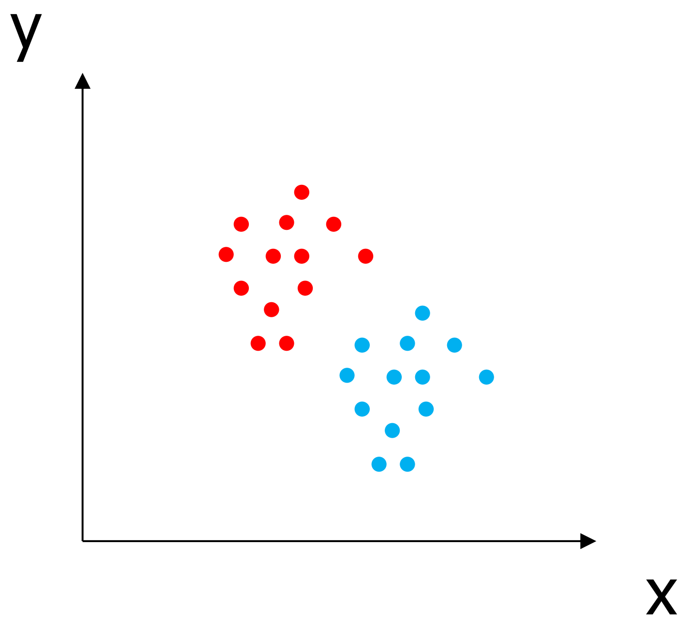
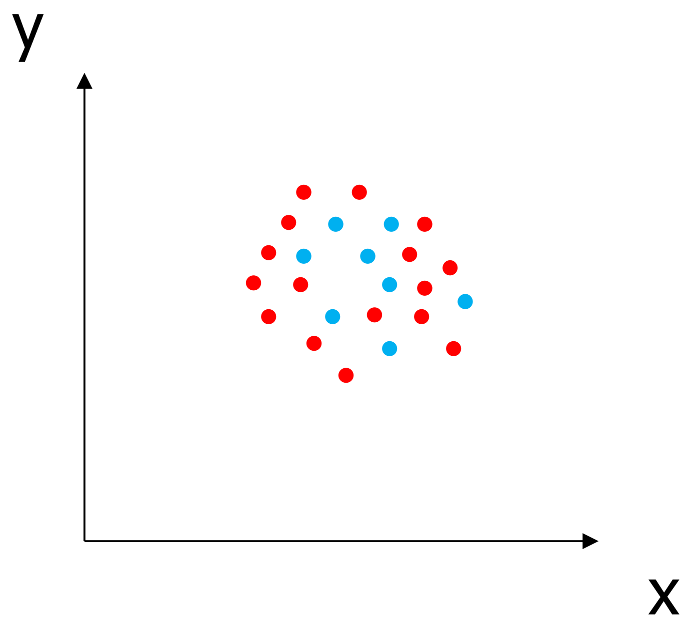
●机器学习常用算法-感知器分类算法



线性可分



线性不可分

●机器学习常用算法-感知器分类算法

感知机目标是找到一个线性可分的超平面：

$$w_1x_1 + w_2x_2 + \cdots + w_mx_m + b = 0$$

感知机

$$f(x) = sign\left(w^T x + b\right)$$

其中输入样本的空间 $X \subseteq \Re^n$ ，输出样本空间$Y = (+1, -1)$，

输入样本 $x \in X$ ，输出样本 $y \in Y$，w 和b是感知模型参数，

符号函数为

$$sign(x) = \begin{cases} +1\ ,\ x \geq 0 \\ -1\ ,\ x < 0 \end{cases}$$

●机器学习常用算法-感知器分类算法

对于一个正样本 x_1 ，若

$$w^T x_1 + b \geq 0$$

则被分为+1类

若负样本 x_1

$$w^T x_1 + b < 0$$

则被分为-1类

●机器学习常用算法-感知器分类算法

感知器模型损失函数

把满足 $w^T x_1 + b \geq 0$ 输出为+1，满足 $w^T x_1 + b < 0$ 输出为-1。

因此正确分类样本满足 $y\left(w^T x + b\right) > 0$

错误分类样本满足 $y\left(w^T x + b\right) < 0$

感知器模型损失函数

对于一个错分样本 x_i ，有 $y_i\left(w^T x_i + b\right) < 0$

该样本到超平面（分类面）的距离是：

$$y_i\left(w^T x_i + b\right) / \|w\|_2$$

令 $\|w\|_2 = 1$

所有错分点到超平面的距离之和为

$$-\sum_{x_i \in E} y_i\left(w^T x_i + b\right)$$

感知器模型损失函数

最终的损失函数为

$$L(w,b)=-\sum_{x_i\in E} y_i\left(w^T x_i+b\right)$$

损失函数的梯度为

$$\nabla_w L(w,b)=-\sum_{x_i\in E} y_i x_i$$

$$\nabla_b L(w,b)=-\sum_{x_i\in E} y_i$$

●机器学习常用算法-感知器分类算法

感知器参数更新

对于一个错误分类点,(x_i, y_i)其对 w,b 的更新如下：

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

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

通过参数更新，不断调整分类面，使损失函数不断减小，最终使其对错误分类点，能够正确分类。

●机器学习常用算法-感知器分类算法

算法：感知器分类算法的原始形式

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

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

（1）选取初值 w_0,b_0

（2）在训练集中选取数据 (x_i,y_i)

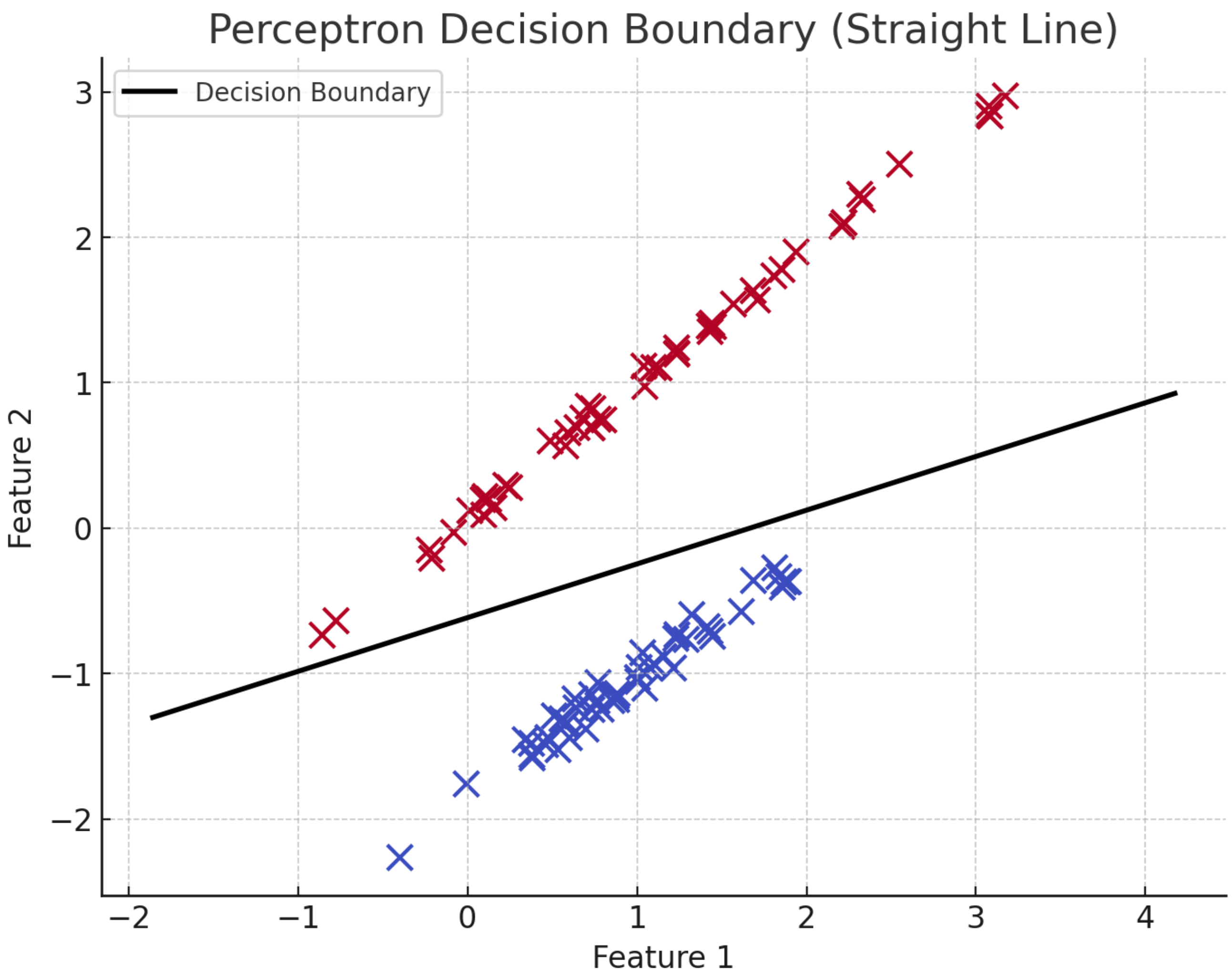
（3）如果 $y_i(w\cdot x_i+b)\le 0$

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

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

（4）转至（2），直至训练集中没有误分类点。

感知机算法在一个二分类数据集上的分类结果如下。图中展示了数据点和由感知机学习得到的决策边界。不同颜色的点代表不同的类别，而决策边界的直线将数据空间划分为两部分。



感知机算法只能找到线性可分数据的决策边界。因此感知机通常会在非线性可分数据上表现不佳，并无法收敛。有其他算法（例如支持向量机或神经网络）可以处理线性不可分的数据。