感知机是一种二分类的线性模型。输入为数据的特征向量，输出为对应的类别，在感知机中取+1和-1。感知机通过寻找一个超平面，将特征空间进行先行划分，正确分离为正负两类。通过导入基于误分类的损失函数，利用梯度下降来最小化损失函数，从而求得感知机模型。利用感知机模型对新的输入数据进行预测分类。感知机算法简单、易于实现，是神经网络和支持向量机的基础。

# 1.感知机模型

输入:x,特征向量

输出:{+1,-1},类别

由输入到输出的函数如下：

 （1）

上式称为感知机,其中，w,b为感知机模型参数，也就是我们待求的参数。通常称w为权重，b为偏置。表示内积。Sign表示:

 （2）

感知机模型的假设空间是定在特征空间上的所有线性分类模型，即函数集合

感知机的集合解释如下：线性方程

 (3)

对于特征空间中的一个超平面S,其中w是超平面的法向量，b是超平面的截距。超平面将特征空间划分为两部分，分别为正类和负类。因此S称为分离超平面。

感知机的学习根据训练数据集:

T={(x1,y1),(x2,y2),...,(xN,yN)}，yi={+1,-1}

求得感知机模型的参数w,b.然后根据参数对新的输入进行计算预测。

# 2.感知机学习策略

## 2.1 数据集的线性可分

给定数据集:

输入:T={(x1,y1),(x2,y2),...,(xN,yN)}，yi={+1,-1}

输出:yi={+1,-1}

如果存在超平面S:



能够将数据集的正负实例点正确的划分到超平面两侧，则对于yi=+1的实例i,有;

对于yi=-1的实例i,有。则称数据集T为线性可分数据集

## 2.2 学习策略

目标:寻找一个超平面，将训练数据集正确地划分为正负实例点。也就是要定义一个损失函数并将其最小化。

损失函数:误分类点到超平面S的距离之和,任一点到平面的距离为:



其中,是w的L2范数。

对于误分类数据(xi,yi)来说:



因此，误分类点xi到超平面的距离为:



假设误分类点集合为M,则所有误分类点到超平面S的距离之和为:



不考虑,就得到感知机学习的损失函数:



显然，损失函数L(w,b)是非负的，如果没有误分类点，损失函数值为0.而且误分类点越少，离超平面越近，损失函数值就越小。

# 3.学习算法

最小化损失函数:



采用随机梯度下降法求解。损失函数L(w,b)的梯度为:





随机选取一个误分类点(xi,yi)对w,b进行更新:





其中，是学习率。通过迭代不断减小损失函数。

直观解释:当一个实例点被误分类，即位于分离超平面的错误一侧时，调整w,b的值，使分离超平面向误分类点的一侧移动，以减小误分类点与超平面间的距离，直至超平面越过该误分类点使其被正确分类。