浅析感知机学习算法

原创 石头 机器学习算法那些事 2018-11-15

前言

核函数和感知机学习算法是支持向量机的基础，支持向量机通过核函数进行非线性分类（参考《深入浅出核函数》），支持向量机也是感知机算法的延伸，本文介绍了感知机学习算法。

目录

1. 感知机模型

感知机是二分类的线性分类模型，由输入特征 **x** 得到输出类别1或-1的映射函数：

$$f(x) = sign(\vec{w}^T \vec{x} + b)$$

称为感知机。其中w，b为感知机模型参数，w为超平面的法向量，b为超平面的截距。若参数确定，则分类模型也相应的确定。sign是符号函数，即：

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

对于新的输入特征 x，分类准则：

$$(1)\text{若}\vec{w}^T \vec{x} + b \geq 0\text{，则结果为正类}$$

$$(2)\text{若}\vec{w}^T \vec{x} + b < 0\text{，则结果为负类}$$

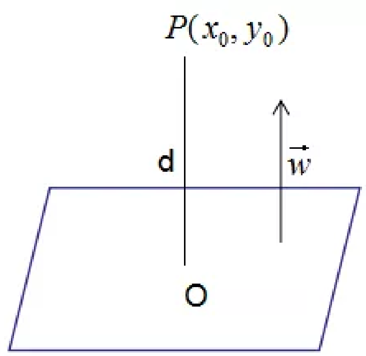
2. 感知机学习策略

1. 点到平面的距离

假设超平面方程为:

$$\vec{w}^T\vec{x}+b=0$$

点$P(\vec{x_0}, y_0)$到超平面的距离d:



$$d=\frac{|\vec{w}^T\vec{x_0}+b|}{\|w\|}$$

2. 感知机学习策略

对于误分类数据(x_i, y_i)，满足如下不等式:

$$-y_i(w\cdot x_i+b)>0$$

成立，因为当$w\cdot x_i+b>0$时，$y_i=-1$，$w\cdot x_i+b<0$时，$y_i=+1$.

因此，误分类点x_i到超平面的距离:

$$-\frac{1}{\|\mathrm{w}\|}y_i(w\cdot x_i+b)$$

正确分类的数据无损失函数，所有误分类的数据点Mi到超平面的总距离为:

$$-\frac{1}{\|\mathrm{w}\|}\sum_{x_i\in M}y_i(w\cdot x_i+b)$$

不考虑标准化范数$\|\mathrm{w}\|$，就得到感知机学习的损失函数:

$$\mathrm{L(w,b)}=-\sum_{x_i\in M}y_i(w\cdot x_i+b) \qquad (2.1)$$

一般用当前样本估计损失函数称为经验风险函数，因此上式就是感知机学习的经验风险函数。

3. 感知机学习算法

2.1式是训练样本的损失函数，显然，损失函数L(w,b)是非负的，在负分类时，损失函数L(w,b)是w,b的连续可导函数，正确分类时，损失函数是0，因此，2.1式是w,b的连续可导函数，可以放心大胆的用随机梯度下降算法来构建模型，梯度下降的方向是损失函数值减小最快的方向，当损失函数为0时，模型构建完成。本节介绍感知机学习算法的两种形式：原始形式和对偶形式。

1. 感知机学习算法的原始形式

损失函数L(w,b)的梯度：

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

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

随机（随机梯度下降法的定义）选取一个误分类点(xi,yi)，对w,b进行更新：

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

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

其中$\eta(0 \le \eta \le 1)$称为学习率

因此，感知机学习算法原始形式的模型构建步骤：

(1) 初始化参数w_0, b_0

(2) 随机选取训练数据(x_i, y_i)

(3)若$y_i(w \cdot x_i + b) \le 0$，则参数更新：

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

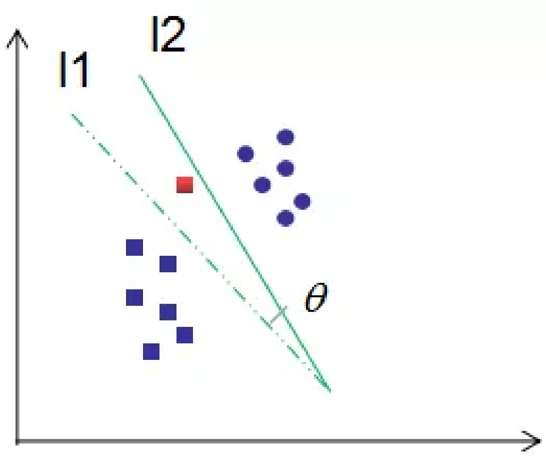
$$b \leftarrow b + \eta y_i$$

(4)转至(2)，直至训练集中没有误分类点

注意：初始化参数w0，b0值不同或随机选取的误分类点不同，得到的最优模型参数也可能不同，因为满足感知机损失函数为0的模型不止一个。

图解感知机学习算法的原始形式：

如下图，红色框为误分类点，l1为位分类直线，随机梯度下降法使l1直线顺时针旋转θ角度为l2直线，点到分类直线的距离逐渐减小直到被正确分类。



2. 感知机学习算法的对偶形式

思想：对于每一个误分类样本点（xi,yi），假设误分类点共迭代n_i次后，结果无误分类点，那么参数w,b就是对应的模型参数，令$\alpha_i = n_i \eta$，参数表示：

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad (3.1)$$

$$b = \sum_{i=1}^{N} \alpha_i y_i \qquad (3.2)$$

解法：用3.1式和3.2世代入上一节的w和b式子，其他步骤完全一样，即可解得参数。

同时，用3.1式和3.2式代入2.1式，可得感知机的对偶模式：

$$f(x) = sign(\sum_{j=1}^{N} \alpha_j y_j x_j \cdot x + b)$$

发现亮点没？f(x)表达式包含了内积部分$x_j \cdot x$，所以尽情的用核函数吧！因此，感知机也能实现非线性分类。

4. 总结

感知机算法有两个点需要引起重视：（1）感知机算法用点到平面的距离作为损失函数，稍微修改下就和支持向量机一样。（2）感知机算法可以写成对偶形式，所以也能通过核函数实现非线性分类。

参考

李航 《统计学习方法》

推荐阅读文章

深入浅出核函数

线性分类模型（一）：线性判别模型分析