

逻辑回归分类器

该课程主要为大家讲授如下的内容：

- 用线性回归解决二分类问题
- 逻辑回归用于二分类

1. 二分类问题

二分类问题是所有分类问题中最简单的，把输入数据划分到两个候选集合中的一个。例如，判断图片中的动物是不是狗。二分类问题也是多分类问题的基础，在大多数情况下多分类问题也可以转化成二分类问题去解决。比如判断图片中的动物是什么？候选的类别包含 鸡，鸭，鹅，猪等。这样一个N分类任务可以转换成N的二分类任务。

1. 用线性回归解决二分类问题

最直接的方法是为模型增加一个阈值. 将输出与阈值比较进行类别划分.

2. 人为设定一个阈值 (threshold) , 比如0

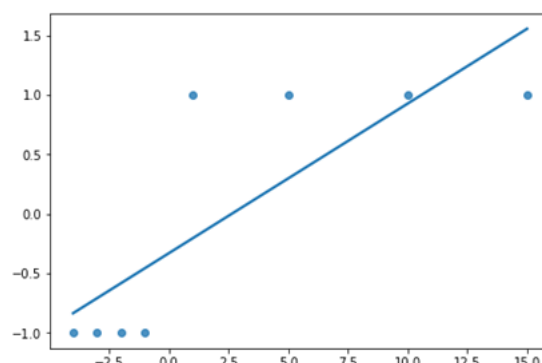
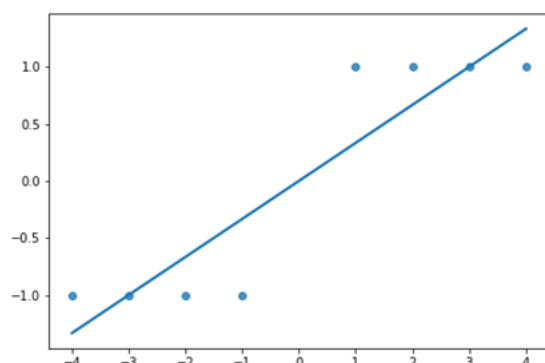
3. 数据预处理, 把是狗的标签值设为1,不是狗设为-1,

4. 模型训练

5. 模型预测

模型输出 $y > 0$ 时，预测图片中是狗；当 $y \leq 0$ 时，预测图片中不是狗。

缺点：模型优化目标和任务目标不一致。任务目标是能正确分类即可(即输出 大于或小于 阈值),但模型优化目标却是输出 = 1 或 -1。



6. 逻辑回归用于二分类

逻辑回归和线性回归有许多相似之处,比如模型目标都是拟合一个

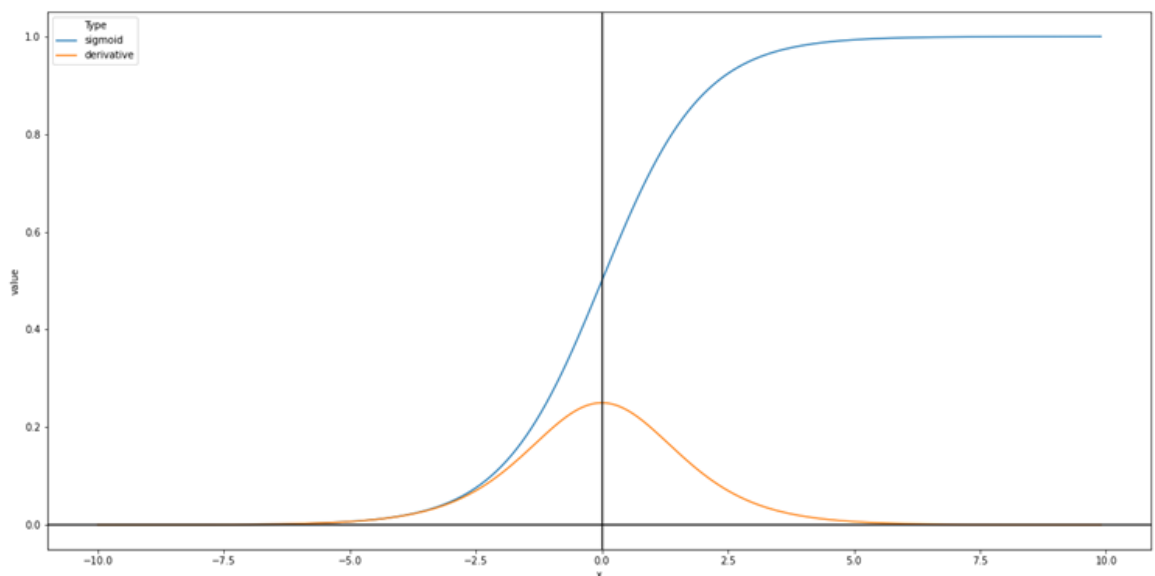
函数,函数输出值为连续值, 不过他们的模型函数和损失函数是不同的。

逻辑回归函数

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Sigmoid函数

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



$\theta^T x$ 表示线性函数的向量化表示方式

所以逻辑回归=线性回归函数+ sigmoid函数

逻辑回归分类器: 模型的输出是属于类别1的概率, 1-模型输出是属于类别0的概率。

事实上用线性回归的损失函数, 也就是最小平方误差, 也是一个有效的学习目标。

但是却没有在逻辑回归中使用它, 原因是线性回归的损失函数在逻辑回归中是一个非凸函数, 难以优化。所以需要构造一个新的损失函数, 这个损失函数需要有两个特点, 第一个是损失函数需要能够和任务的目标一致, 也就是说当任务准确度高时损失函数应该小, 反之亦然。

第二是损失函数要容易优化, 最好是一个凸函数。

整合后的损失函数：

$$J(\theta) = - \sum_{i=1}^m [y^i \cdot \log h_{\theta}(x^i) + (1 - y^i) \cdot \log (1 - h_{\theta}(x^i))]$$