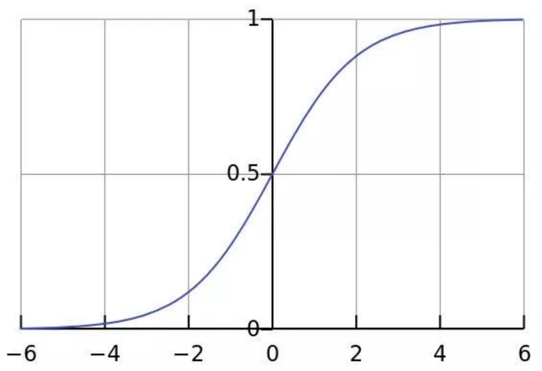
Logistic Regression 逻辑回归

sigmoid 函数

相较于线性回归的因变量 y 为连续值，逻辑回归的因变量则是一个 0/1 的二分类值，这就需要我们建立一种映射将原先的实值转化为 0/1 值。这时候就要请出我们熟悉的 sigmoid 函数了：

$$y = \frac{1}{1+e^{-z}}$$

其函数图形如下：



该函数具有很好的数学性质，既可以用于预测类别，并且任意阶可微，因此可用于求解最优解。将函数带进去，可得 LR 模型为

$$y = \frac{1}{1+e^{-(w^T x+b)}}$$

sigmoid 函数还有一个很好的特性就是其求导计算等于下式，这给我们后续求交叉熵损失的梯度时提供了很大便利。sigmoid 函数求导：

$$y' = y \times (1-y)$$

其实，LR 模型就是在拟合 z = w^T x +b 这条直线，使得这条直线尽可能地将原始数据中的两个类别正确的划分开。

损失函数

LR 解决二分类问题中，交叉熵损失函数为如下形式

$$L = -\left[y \log \hat{y} + (1-y) \log(1-\hat{y})\right]$$

这个函数通常又称为对数损失 logloss，这里的对数底为自然对数 e ，其中真实值 y 是有 0/1 两种情况，而推测值 y^ 由于借助对数几率函数，其输出是介于 0~1 之间连续概率值。因此损失函数可以转换为分段函数

$$L = \begin{cases} -\log(\hat{y}), & y = 1 \\ -\log(1-\hat{y}), & y = 0 \end{cases}$$

优化求解

确定损失函数后，要不断优化模型。LR 的学习任务转化为数学的优化形式为

$$(w^*, b^*) = \arg\min_{w,b} L(w,b)$$

是一个关于 w 和 b 的函数。同样，采用梯度下降法进行求解，过程需要链式求导法则

$$\frac{\partial L}{\partial W} = \frac{1}{m} x(\hat{y} - y)$$

$$\frac{\partial L}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)$$

$$w \leftarrow w - \alpha \frac{\partial L}{\partial w}$$

$$b \leftarrow b - \alpha \frac{\partial L}{\partial b}$$