logistics回归

上一页笔记提到两种模型：生成模型与鉴别模型

生成模型：假设每个类别有一个专属语言模型，将文本d放入每个模型中跑，找到生成该文本概率最大的模型，输出它的类别C

核心动作：计算 C，d 联合概率 $P(c,d)$

核心公式：$C_{MAP} = \underset{c \in C}{\arg\max}\ P(d|c)P(c)$

鉴别模型：直接计算 $P(c|d)$

二元Logistics回归：典型的鉴别模型

给定输入输出训练集 $(x^{(i)}, y^{(i)})$ i表示第i个样本，$x^{(i)}$ 表示这个样本的输入特征（如特征向量），$y^{(i)}$ 表示其真实类别，在二元逻辑回归里y只有两个值：0（负例），1（正例）

对于每个 $x^{(i)}$，模型会先将其处理为特征向量 $[x_1, \cdots, x_n]$ 然后计算输出预测类别 $\hat{y}^{(i)}$

具体计算：逻辑回归是一个线性模型，会先计算一个线性得分：

$$z^{(i)} = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$
$$= w \cdot x^{(i)} + b$$

$w = [w_1, \cdots, w_n]$ 是给每个特征分配一个权重，表示该特征的重要程度

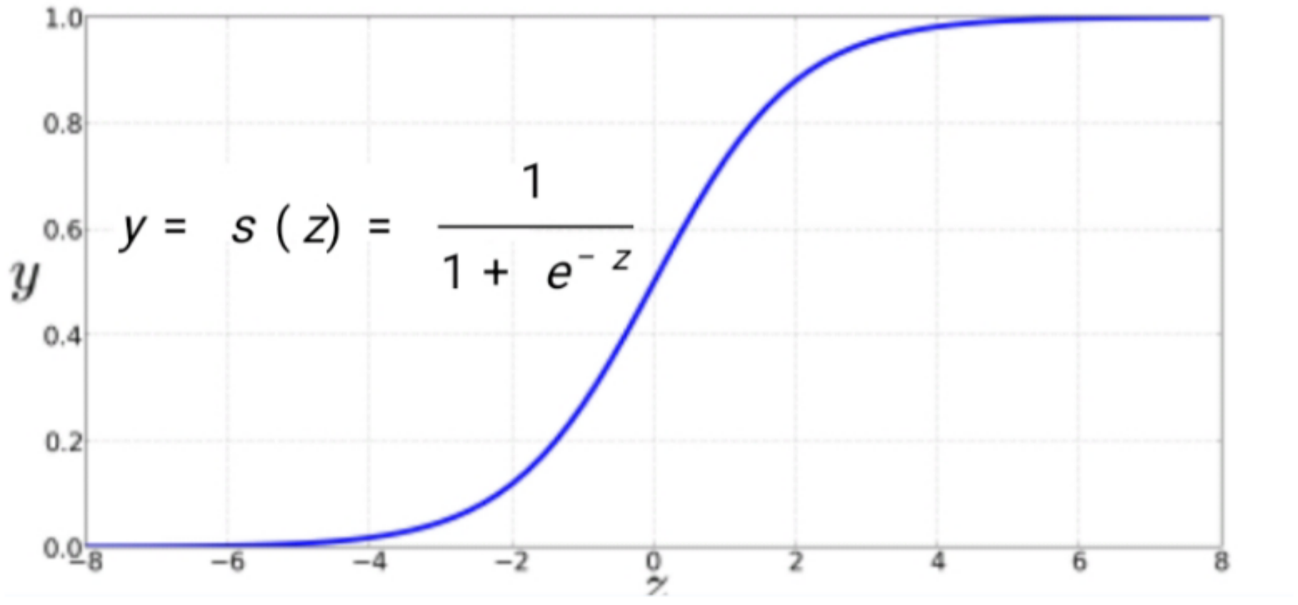
b是偏置（截距），z是线性加权和

如果z大于规定值，则 $y=1$，否则为0

归一化：Sigmoid激活函数，将z转化为概率，更利于将其用于做决策

$$y = S(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+\exp(-z)}$$

$y \to 1$ 表示模型极度确信是正类



S(z)也写作 $\sigma(w \cdot x + b)$

Sigmoid函数计算概率

正类 $y=1$ 的概率：$P(y=1|x) = \frac{1}{1+\exp(-(w \cdot x + b))}$

负类 $y=0$ 的概率：$P(y=0|x) = 1 - P(y=1|x)$

$= \frac{1}{1+\exp(w \cdot x + b)}$

决策：$\hat{y} = \begin{cases} 1 & \text{if } P(y=1|x) > 0.5 \\ 0 & \text{otherwise} \end{cases}$

0.5被称为决策边界

Logistics回归学习

模型：观测特征向量：x

类别集合：C

输出：$\hat{y} = \sigma(w \cdot x + b)$

真实类别：y

参数学习：设置参数w和b最小化 $\hat{y}^{(i)}$ 与 $y^{(i)}$ 的距离（损失，$L(\hat{y}, y)$）

似然概率：$P(y|x) = \hat{y}^{y} \cdot (1-\hat{y})^{1-y}$

这是单个样本的似然函数，衡量的是预测结果与

真实类别的匹配程度，值越接近1说明模型越准

当 $y=1$ $P(y|x)=\hat{y}$ → 模型预测为正类的概率

当 $y=0$ $P(y|x)=1-\hat{y}$ → 模型预测为负类的概率

我们的目的：最大化 $\log P(y|x)$

$$\log P(y|x) = \log\left[\hat{y}^{y}(1-\hat{y})^{1-y}\right]$$

$$= y\log\hat{y} + (1-y)\log(1-\hat{y})$$

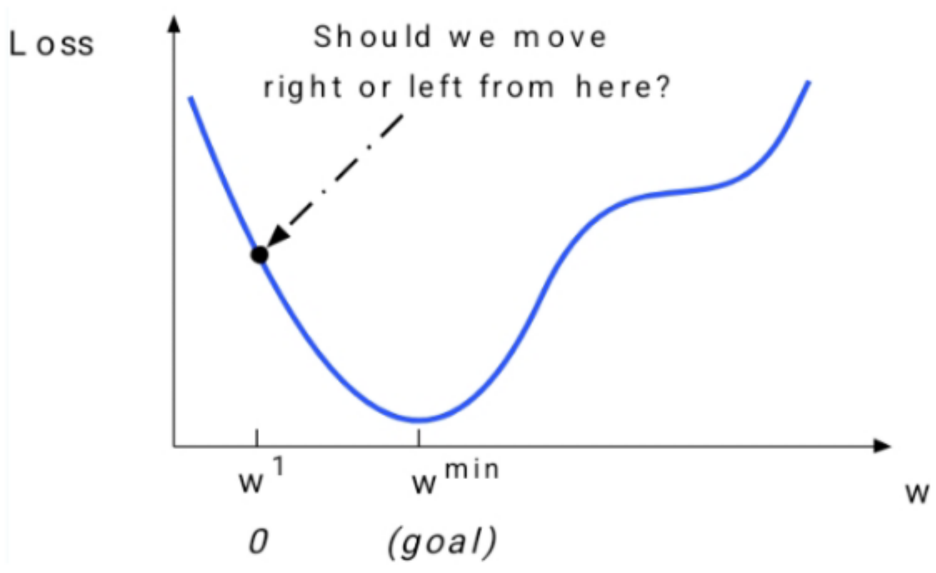
交叉熵损失：$L_{CE}(\hat{y},y) = -\log P(y|x) = -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$

最大化对数似然 → 最小化交叉熵损失

梯度下降：参数集合表示为 $\theta=(w,b)$，$\hat{y}=f(x;\theta)$

最小化参数使所有样本平均损失最小化：

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{m}\sum_{i=1}^{m} L_{CE}(f(x^{(i)};\theta), y^{(i)})$$



$$w^{t+1} = w^{t} - \eta\frac{d}{dw}\hat{\theta}$$

梯度：$\frac{d}{dw}L(f(x;w),y)$

步长：η

随机梯度下降（SGD）：每次迭代随机选1个样本计算梯度而非用全量样本（用全量样本叫批量梯度下降，用部分样本的叫小批量梯度下降）