

Logistic函数和Softmax函数

这里简单总结一下机器学习最常见的两个函数，一个是 `logistic` 函数，另一个是 `softmax` 函数。首先分别介绍 `logistic` 函数和 `softmax` 函数的定义和应用，然后针对两者的联系和区别进行了总结。

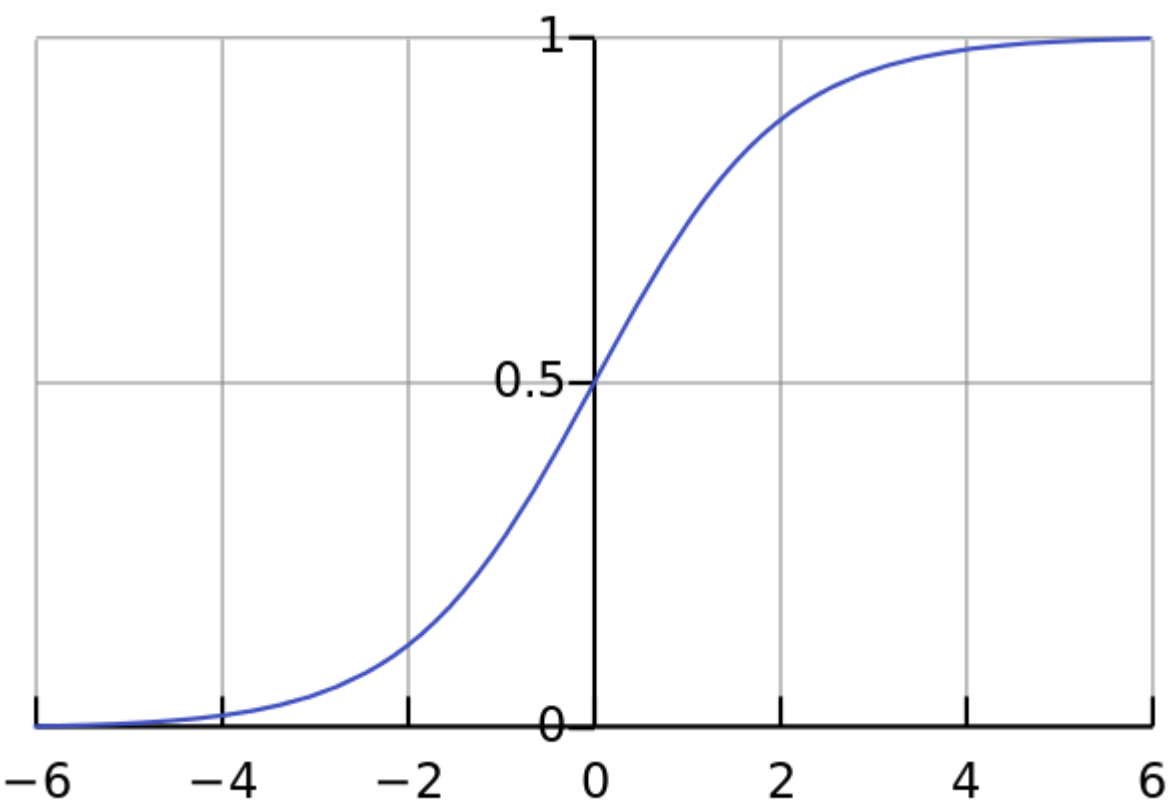
logistic函数

logistic函数的定义

logistic 函数也就是经常说的 *sigmoid* 函数，它的几何形状也就是一条 *sigmoid* 曲线。*logistic* 函数的公式形式如下：

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

其中 x_0 表示函数曲线的中心 (`sigmoid midpoint`)， k 是曲线的坡度，函数的形状如下图。



实际应用

logistic 函数本身在众多领域中都有很多应用，这里只谈统计学和机器学习领域。*logistic* 函数在统计学和机器学习领域应用最为广泛或者最为人熟知的肯定是逻辑斯谛回归模型。逻辑斯谛回归 (`Logistic Regression`，简称 `LR`) 作为一种对数线性模型 (`log-linear model`) 被广泛地应用于分类和回归场景中。此外，*logistic* 函数也是神经网络最为常用的激活函数，即 *sigmoid* 函数。

Softmax函数

定义

softmax is a generalization of logistic function that "squashes"(maps) a K -dimensional vector z of arbitrary real values to a K -dimensional vector $\sigma(z)$ real values in the range $(0, 1)$ that add up to 1.

上述表面了 **softmax** 函数与 **logistic** 函数的关系，也阐述 **softmax** 函数的本质就是将一个 K 维的任意实数向量压缩（映射）成另一个 K 维的实数向量，其中向量中的每个元素取值都介于 $(0, 1)$ 之间。

$$\sigma(z) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

其中 $j = 1, 2, \dots, K$

实际应用

softmax 函数经常用在神经网络的最后一层，作为输出层，进行多分类。此外 **softmax** 在增强学习领域内，**softmax** 经常被用作将某个值转化为激活概率，这类情况下 **softmax** 的公式如下：

$$P_t(a) = \frac{e^{\frac{q_t(a)}{T}}}{\sum_{i=1}^n e^{\frac{q_t(i)}{T}}}$$

其中 T 被称为是温度参数（temperature parameter）。当 T 很大时，即趋于正无穷时，所有的激活值对应的激活概率趋近于相同（激活概率差异性较小）；而当 T 很低时，即趋于 0 时，不同的激活值对应的激活概率差异也就越大。

两者之间的关系

1. **logistic** 函数具体针对的是二分类问题，而 **softmax** 函数针对的是多分类问题，因此可以将 **logistic** 函数看作是 **softmax** 函数的一个特例。

参考 [UFDL](#) 中 **softmax** 回归的推导，如下：

当分类数目为 2 时，**softmax** 回归的假设函数表示如下：

$$h_{\theta}(x) = \frac{1}{e^{\theta_1^T x} + e^{\theta_2^T x(i)}} \begin{bmatrix} e^{\theta_1^T x} \\ e^{\theta_2^T x} \end{bmatrix}$$

利用 **softmax** 回归参数冗余的特点，从两个参数向量中都减去向量 θ_1 得到：

$$\begin{aligned} h_{\theta}(x) &= \frac{1}{e^{\vec{0}^T x} + e^{(\theta_2 - \theta_1)^T x(i)}} \begin{bmatrix} e^{\vec{0}^T x} \\ e^{(\theta_2 - \theta_1)^T x} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{1 + e^{(\theta_2 - \theta_1)^T x(i)}} \\ \frac{e^{(\theta_2 - \theta_1)^T x}}{1 + e^{(\theta_2 - \theta_1)^T x(i)}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{1 + e^{(\theta_2 - \theta_1)^T x(i)}} \\ 1 - \frac{1}{1 + e^{(\theta_2 - \theta_1)^T x(i)}} \end{bmatrix} \end{aligned}$$

设定 $\theta' = \theta_2 - \theta_1$ ，则上述公式可以表述为 **softmax** 回归器预测其中一个类别的概率为：

$$\frac{1}{1 + e^{(\theta')^T x(i)}}$$

则另一个类的概率为：

$$1 - \frac{1}{1 + e^{(\theta')^T x(i)}}$$

可以看到这里与*logistic*回归是一致的

- 从概率角度来看两者的区别。*softmax*建模使用的分布是多项式分布，而*logistic*则基于伯努利分布，这方面的解释可以参考 *Andrew Ng* 的讲义。
- softmax*回归和多个*logistic*回归的关系

多个*logistic*回归通过叠加也同样可以实现多分类的效果，那么多个*logistic*回归是不是和*softmax*一样呢？*softmax*回归进行的多分类，类与类之间是互斥的，即一个输入只能被归为一类；而多个*logistic*回归进行多分类，输出的类别并不是互斥的，如"苹果"这个词语既属于"水果"类也属于"3C"类别。

参考链接

[1. Logistic Function Wiki](#)

[2. Softmax Function Wiki](#)

[3. Softmax回归](#)