# 矩阵求导



以下用小写字母代表标量，例如；小写黑体代表向量，例如；大写黑体代表矩阵，例如。为所有n行m列实数矩阵(有时为了令维数更醒目也记作)。

## 标量对矩阵求导、矩阵对标量求导

**标量对矩阵(向量)求导、矩阵(向量)对标量求导，求导后结果与原矩阵(向量)同型。**

标量对矩阵(向量)求导：

若函数，，则



矩阵(向量)对标量求导：

若函数，，则



## 向量对向量求导

若函数，，则



## 雅各布(Jacobian)矩阵

设函数，，则对的雅各布矩阵定义如下：



其中：



## 一些常用矩阵求导公式的推导



所以





# BP算法

## 神经网络记号

设神经网络层数为层，编号为1至，其中含有1个输入层(记为层1)、个隐层(记为层2~层)，1个输出层(记为层)。层也记作，为层的结点数量，为层的结点。层的**输入**向量记为



其中代表层i的j结点()的输入，**输出**向量记为



输入层(层1)的，其中为输入。

层i与层i+1的结点之间的**权重**矩阵记为



其中：为结点到的权重。

其余各层的输入、输出、权重之间的关系如下：







其中，称为sigmoid激活函数。

神经网络的输出



神经网络输入向量的标签向量记作

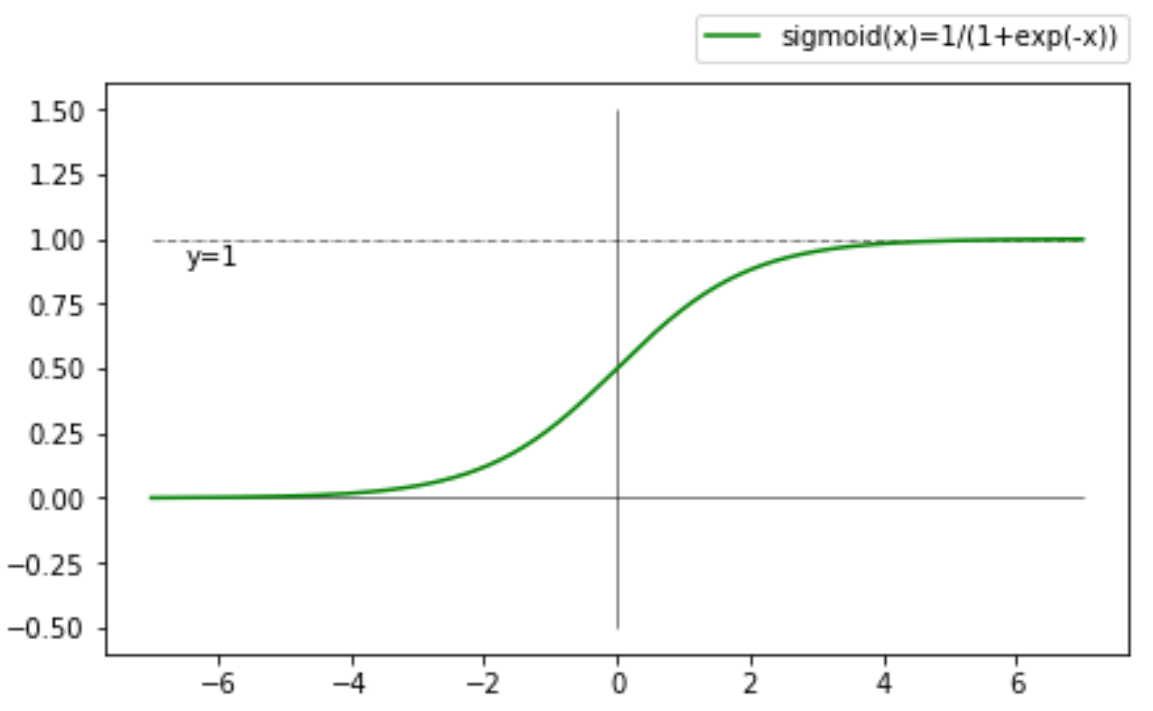
## 激活函数

### sigmoid激活函数



其导数为：





图表 1：sigmoid激活函数

## 损失函数

损失函数：



注意：根据损失函数的定义，此处需要计算与的内积(损失函数的值是标量)，因此使用转置形式。

## 损失函数对权重求导

为了使用梯度下降方法更新权重，需要计算损失函数(简记为)对的偏导数：



(注意是的函数，而是的函数。而当时，不是的函数，因为不影响。)

定义误差因子如下：



其中：



则有



由上式求出的矩阵元素倒推出矩阵公式如下：



## 计算误差因子



其中：，f为sigmoid激活函数。

因此



矩阵形式如下：



由可知，因此上式



其中表示两个同维度向量对应元素相乘，得到一个同维度的新向量，例如



因为：，有

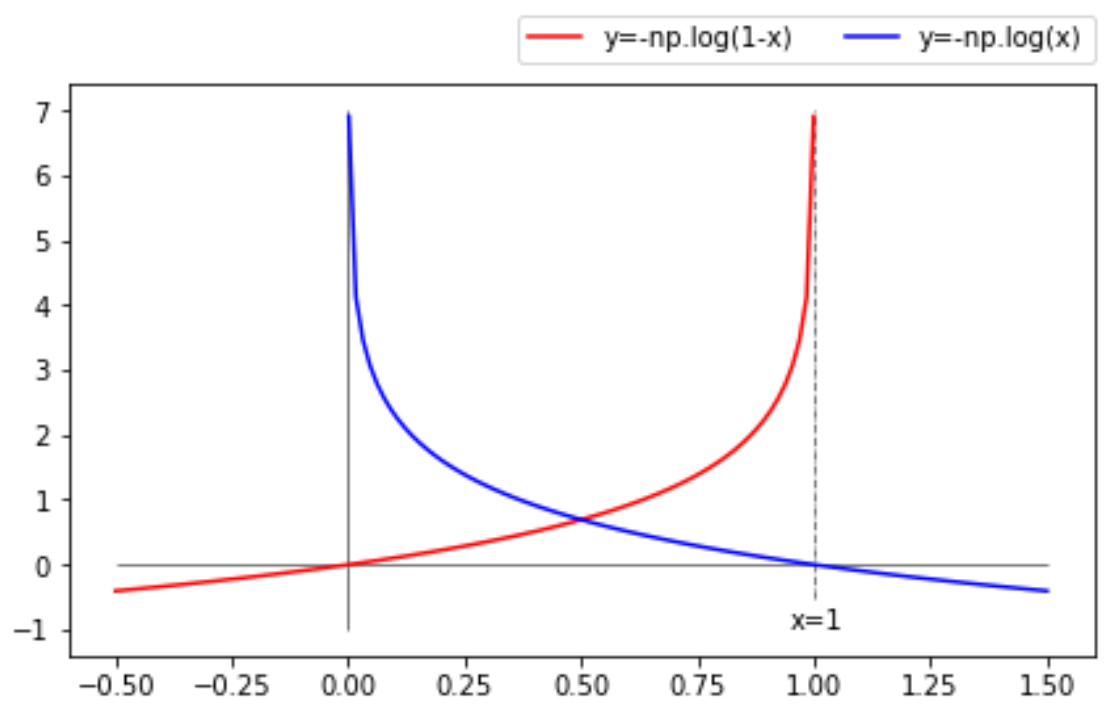


因此由元素推导出相应矩阵公式如下：



# 交叉熵损失函数原理





图表 2：交叉熵损失函数原理图(当时，如果趋于1，或者当y=1时，如果趋于0，则为无限大)