# 数学铺垫

## 伯努利分布

定义 随机变量服从**伯努利分布**，如果只能取0、1两个值，且取这两个值的概率分别为

伯努利分布的分布函数为

## Sigmoid函数

定义 Sigmoid函数

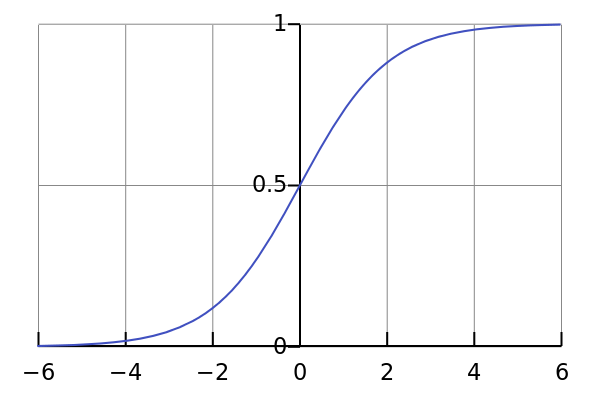


图 1 Sigmoid函数图像

性质 导数可以用自己表示

## 维数灾难（Curse of Dimensionality）

为了抑制维数灾难，我们需要对数据的“天性”——数据服从何种分布，做出合理的假设。

## 最大似然估计（Maximum Likelihood Estimation）

Let the rv’s form a random sample from a discrete or a continuous distribution whose p.f. or p.d.f. is .

For every observed vector in the sample, the value of the **joint** p.f. or the joint p.d.f. is denoted as .

Generally, random variables from a random sample is considered to be independent. Thus, we have

To simplify optimization works, we yield the **log-likelihood function**, which is also called cross-entropy loss function.

Our optimization goal is

借用某博客中的一句话。**对于一个模型已定、样本已有、参数未知的模型（这就是statistical learning），MLE就是找出一组参数，使得模型产生出样本的概率最大。**

# 线性回归

定义 线性回归假定输出是输入的**线性变换**，也就是

其中，是**权重**向量和**输入**向量之间的**数量积**。是预测值和实际输出之间的**残差**。

同时，我们认为，残差服从正态分布。如此，我们可以用另一种方式来表达线性回归模型：

最简单地，我们可以令，同时令为固定值。这种情况下，模型的**参数**为

例子 假设输入是一维的，我们可以把期望的输出表示成

其中，是截距，或者叫**bias**。这里用到了一个技巧：为输入追加一个维度，值固定为1。也就是.

引申 线性回归可以用来建模非线性关系，只需要把替换成某种非线性变换.

# logistic回归

## 模型介绍

把线性回归的高斯分布换成伯努利分布，再用上Sigmoid核函数，就是

其中

注意 Logistic回归虽然名字里有回归，但**它是一种二分类（Classification）**，而不是回归（Regression）。它叫这个名是因为idea来自线性回归。

## 用极大似然法（MLE）估计参数W

回忆 交叉熵损失函数

逻辑回归的负对数似然函数为