信用分模型

目前有n个样本，每个样本表示一个贷款客户的信息。则第i个样本表示为\{x_i,y_i\}，其中x_i\in R^{r\times1}，y_i\in\{0,1\}。

具体地说，x_i是一个维向量，每一个维度表示贷款人的一个属性变量。y_i的值是0或1，y_i=1表示逾期，y_i=0表示正常。

写成矩阵的形式，n个样本全部数据，可以写成如下形式:



对客户i而言，如果其逾期的概率是p_i，那么其正常的概率就是1-p_i，一个客户要么逾期要么正常，逾期和正常的概率之和必然是1。

因为需要根据x_i评估用户i是否会逾期。所以可以考虑使用最常见的广义可加模型，其中线性模型最为简单。

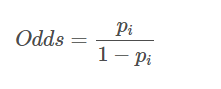
如下所示：:

z_i=\beta_0+\beta_1x_{i,1}+\beta_2x_{i,2}+...+\beta_rx_{i,r}

另外，又考虑到概率p的值域是(0,1)区间，希望输出的函数值在这个值域。因此再嵌套一层函数以变换值域：  
p_i=\frac{e^{z_i}}{1+e^{z_i}}=\frac{1}{1+e^{-z_i}}

显然，上式的值域一定是在(0,1)区间的，这里使用e^x函数是因为e^x在微分求导上更便利。

我们知道，传统的金融风控模型会涉及到Odds（几率）：它指的是事件应该发生的概率与事件不应该发生的概率的比值，在这里可以定义为：



ln\frac{p_i}{1-p_i}=z_i