

如何将类别形变量引入线性回归：

资料查到：

1. one-hot encoding

理解：

三个样本，三个特征,Feature_1 有 ‘大’ ‘小’ Feature_2 有 ‘红’ ‘绿’ ‘蓝’ Feature_3 有 ‘好’ ‘坏’

	Feature_1	Feature_2	Feature_3
Sample_1	大	红	好
Sample_2	小	蓝	好
Sample_3	大	绿	坏

编码后

	Feature_1	Feature_2	Feature_3
Sample_1	10	001	10
Sample_2	01	010	10
Sample_3	10	100	01

则当一个样本为（小，蓝，好）

经过处理后为

（0101010）

2. binary encoding

同样，先将特征转换为数字

再转为二进制

好	1 (0)
坏	2 (10)
蓝	3 (11)
绿	4 (100)
大	5 (101)