# QA

1. 两个类别型变量构造笛卡尔特征组合为什么能提升模型表现

笛卡尔特征组合可以看做是把两个类别型变量穷举出所有可能的组合方式构造出n新的特征，每一个特征的值用“1，0”表示是否同时体现出两个特征。

没有找到相关数学推导的原理（不知道笛卡尔乘积这种运算会不会让某一种组合方式的信息熵变大），感性理解的原因是：笛卡尔乘积可以看做是一种特征组合的方式，可以有助于表示非线性关系；另外笛卡尔乘积进行了特征的离散化，这样做的优势有以下几个：1）特征离散化之后，起到了简化模型的作用，使模型变得更稳定，降低了模型过拟合的风险。2）离散化之后的特征对异常数据有很强的鲁棒性，实际工作中的哪些很难解释的异常数据一般不会做删除处理，如果特征不做离散化，这个异常数据带入模型，会给模型带来很大的干扰。3）离散特征的增加和减少都很容易，且稀疏向量的内积乘法运算速度快，易于模型的快速迭代。4）特征离散化之后，每个离散变量都有单独的权重，相当于给模型引入了非线性，能够提高模型的表达能力。