# 问题

## 特征工程

### 梯度与量纲大小有关吗？

梯度一般而言来自激活函数f(x)’ 和x的累积。而激活函数不同其结果也不同。如sigmoid函数导数为y（1-y）并非单调函数，梯度绝对值仍然会很大（反方向）；tanh导数则为1-tanh^2 是一个值在-1 到1之间的单调增但是快速饱和到1的函数。Relu函数导数1。

### 归一化的作用

当梯度与量纲关系有关，各个维度学习率相同，量纲大的维度步长更大，步长与梯度大概差一阶，在该维度上需要更多迭代数才能找到最优解。

### N-gram

n-gram时空复杂度？

计算概率复杂度大概在k，语料库词数。空间复杂度v的n次方 v为词表大小

N越大越好吗？N越大空间复杂度很高，容易出现稀疏的n-gram组合，模型计算使用连乘，导致结果为0。

平滑是为了做什么？使用平滑，保证最小概率不为0，防止0概率的组合出现影响整个句子的处理。最简单的平滑手段就是+1

# 记忆

## 特征工程

### 归一化的种类

线性函数归一化：平移到0~1区间

零均值归一化：转换为0 1 正态分布。

### n-grams计算

### 文本表示种类

词袋模型：tf-idf tf\*idf tf代表词频 idf代表逆文档频率（词在所有文章中出现的频率）

主题模型：6.5

n-grams：词嵌入