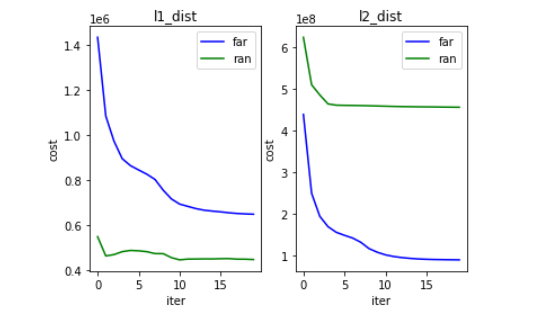
# 1、使用 Spark 实现 k-means

代码



代价函数关于迭代次数的关系绘图



在欧氏距离作为度量下随机数据变化了26.48%，远距离数据变化了76.70%。由于远距离的数据使得簇的分布间隔较远，每个簇也相对比较稳定。

在曼哈顿距离作为度量下随机数据变化了18.65%，远距离数据变化了51.55%。远距离数据是根据欧式距离得出，在曼哈顿距离下不一定相距很远。

# 推荐系统

: 用户i喜欢的商品数量。

: 用户i与用户j共同喜欢的商品数量。

或是为了在求余弦相似度时分母相乘得到1，于是乘上自己的转置就得到余弦相似度。

与相乘，通过用户与用户的相似度计算得到推荐值。

与相乘，通过商品与商品的相似度计算得到推荐值.

用户-用户协同过滤方法下为 Alex 推荐的 5 部相似得分最高的电视节目：

"FOX 28 News at 10pm",

"Family Guy",

"2009 NCAA Basketball Tournament",

"NBC 4 at Eleven",

"Two and a Half Men"

用品-商品协同过滤方法下为 Alex 推荐的 5 部相似得分最高的电视节目：

"FOX 28 News at 10pm",

"Family Guy",

"2009 NCAA Basketball Tournament",

"NBC 4 at Eleven",

"Two and a Half Men"

