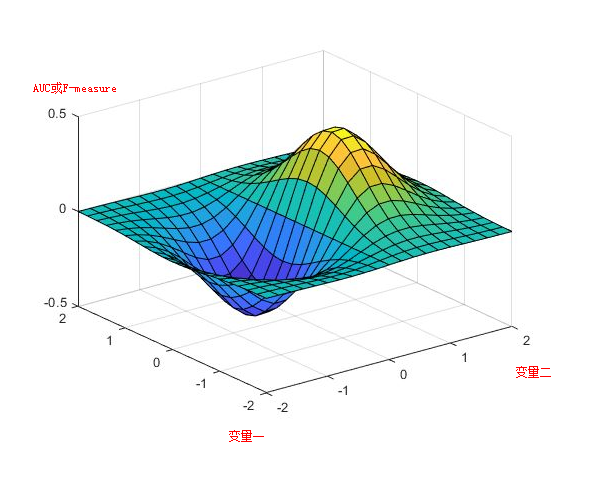
超级码力数据挖掘工程师选拔测试

第一题：

1. 用一个你觉得效果最好的分类算法（e.g., SVM，RF，ANN，etc.）对Test1的data进行classification建模，用5-fold crossvalidation来计算预测的AUC或F-measure。若电脑计算能力不足可用3-fold。对于有些本身就含有cross validation功能的算法可以不用cross validation来稳定结果。（PS，对于非计算机系的同学，以python或R为例，各种分类算法以及AUC的计算都可以在网上找到对应参考代码。）
2. 把你认为这个算法里的最重要的两个参数用网格优化（grid search）的方式进行优化，并把过程画出来。输出参数优化过程的三维图（2个参数变量 + 5 or 3 fold的AUC或者F-measure的值），示意图如下。



1. 提交的作业结果要求包含：
   1. 可直接运行的代码（注意路径）。
   2. 参数优化过程的结果图。
   3. 一个简易说明文档。

第二题：

1. 对Test2的data进行association rule analysis，挖掘的rule要求具备如下条件：
   1. Rule的右侧只有一项，且其为Test2\_data中的最后一列“label”。（eg. A->label, BC -> label, AD -> label）
   2. Rule的左侧少于等于两项。
   3. Support 0.1以上，Confidence 0.7以上。
2. 提交的作业结果要求包含：
   1. 可直接运行的代码（注意路径）。
   2. 一个简易说明文档，其中列明挖出的rule。