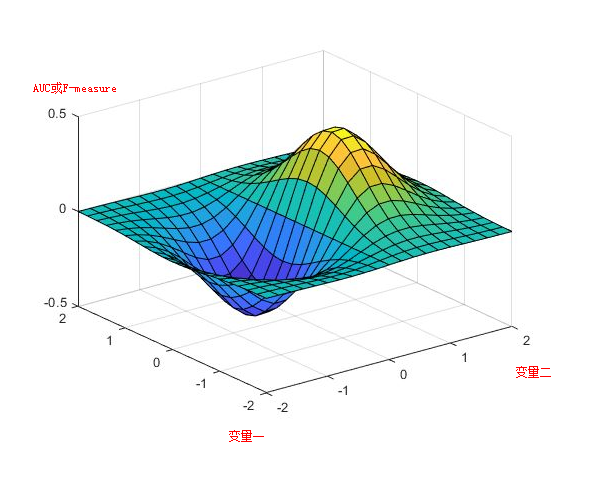
数据挖掘XX生选拔测试

（下列两题中任选一题，无特殊情况请选择第一题）

第一题：

1. 用3种你最喜欢的分类算法（e.g., **SVM，RF，ANN**，etc.）对training 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. 挑其中一种表现最好的算法（另外两种不要求你优化参数），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. 作业包含：1. 三种算法的各自可直接运行代码。2. 其中一种算法的三维的参数优化过程的代码及结果图。
2. 最终作业以三维的优化过程图的美观易理解性、最佳准确率以及全部的代码质量作为评判标准。
3. 语言要求下列三者之一：R、Python（优先）、Matlab。
4. 只要你有python的数据挖掘经验，此题会很容易。完成时间不做要求，完成后就安排面试。

第二题：

如果你不选择第一题，那么我们希望你至少在深度学习或者科研领域有一定经历，否则请做第一题。此题考查你的表达总结能力，你的作业是把你最值得分享的项目制作成ppt让我们了解了解。但我们对ppt有所要求：

1. 希望ppt里不包含任何的技术细节，想象它的读者不是圈内人士，但读者要能明白你技术或者算法的大致逻辑。
2. 整个项目你必须要讲清楚全部的why what how。比如，项目解决了什么问题，意义何在，用了什么方法，此方法独到之处在哪儿，解决的结果如何，有哪些启发意义等。
3. PPT要求图文并茂，清晰易懂。