《计算机应用数学》作业实验报告

李斌 11531041

作业1. 曲线拟合

**实验原理：**

多项式线性回归，给出一组数据*x*，依据题目给定的要求，计算其多项式特征，然后通过线性回归的方式进行求解。

具体地，若设计矩阵为*X*，参数矩阵为*Θ*，目标向量为*Y*，则：*XΘ≈Y*

那么：

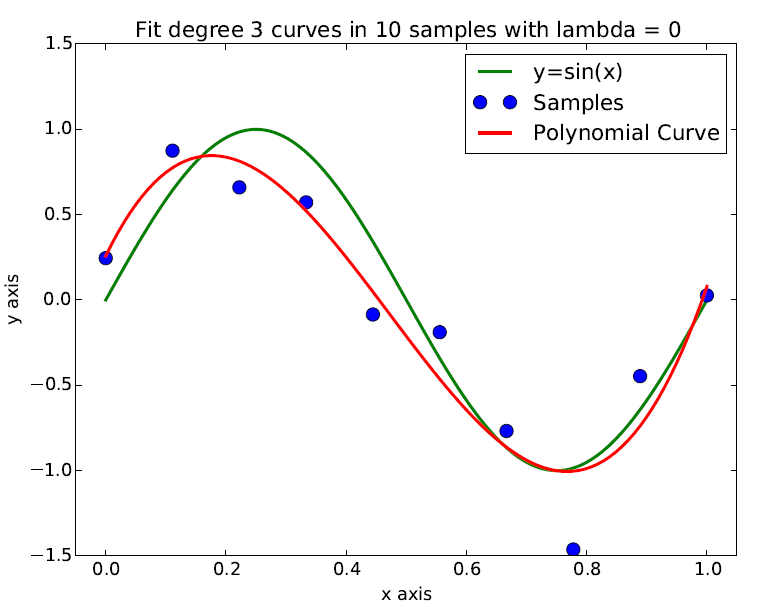
*Θ*\*=(*X*T*X*)-1*X*T*Y*

原始数据的获取：

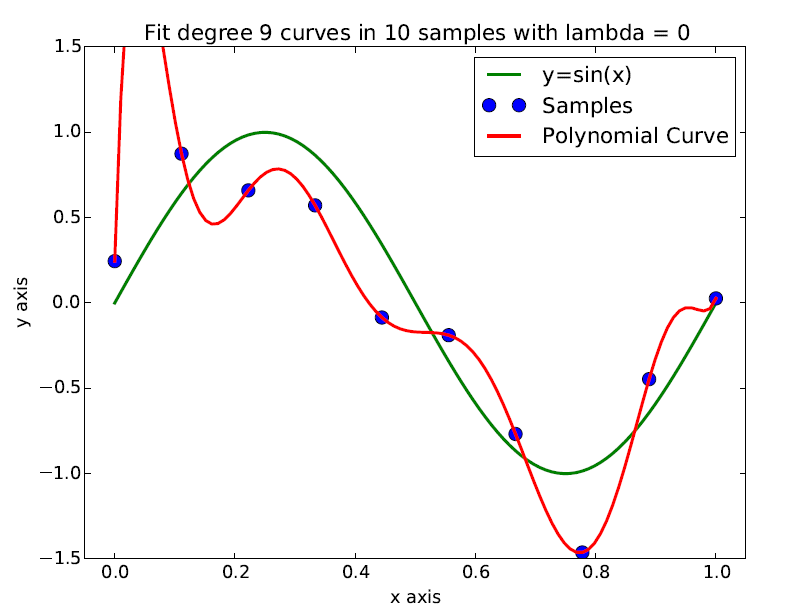
首先，计算*y*=sin*x*的数据向量，并依据题意进行抽样。然后计算高斯分布的白噪声向量。将二者进行叠加即为最终的结果。

**实验效果：（需要说明，参数lamdba为正则化项系数）**

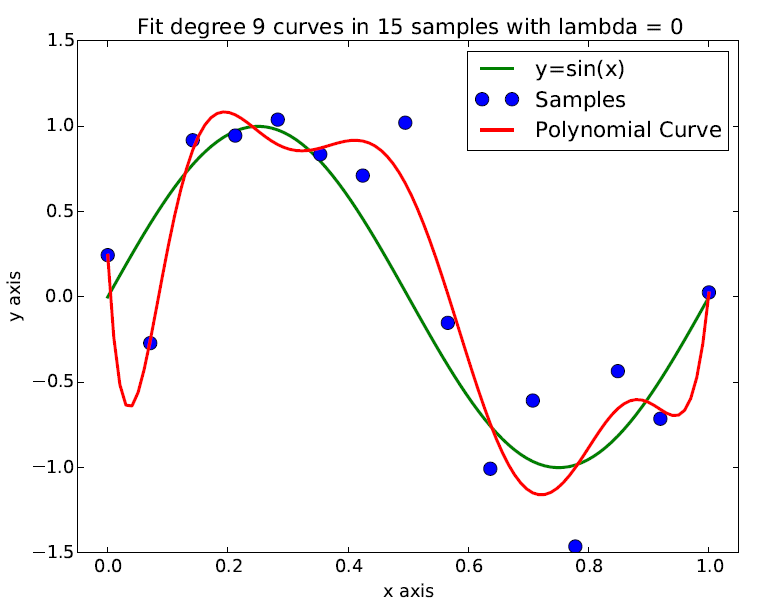
1.



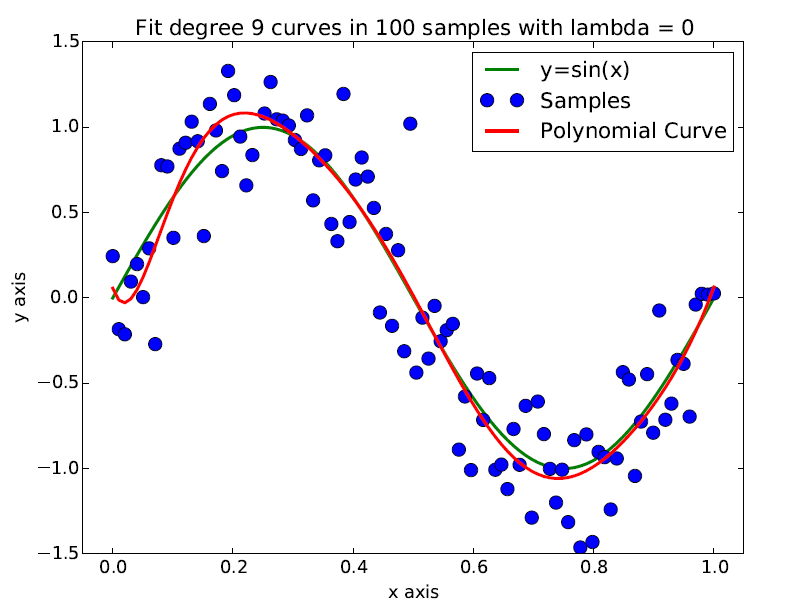
2.



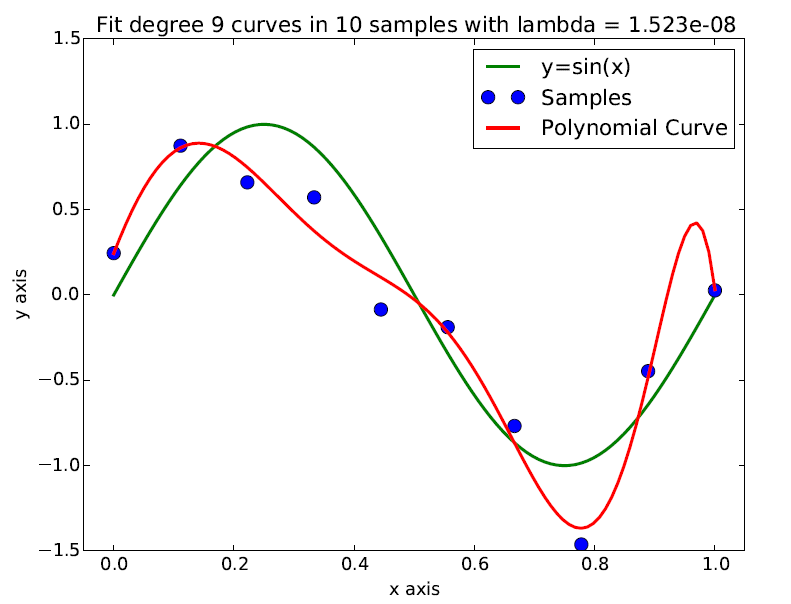
3.



4.



5.



作业2. PCA（主成分分析）

**实验原理：**

主成分分析，即通过特征值理论，将数据分解为特征矩阵的乘积、或者利用奇异值分解，将数据进行分解。然后，提取特征值的前*k*个分量，也即主分量。

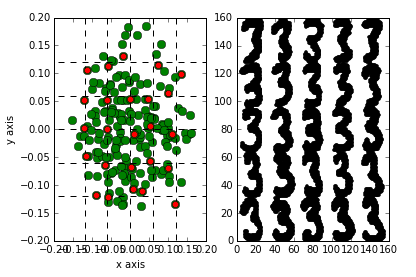
当我们将普通的数据，映射到不同的子空间中去，即可滤除冗余特征。

具体的方案。首先获得UCI数据库中关于3的所有数据，并保存（利用python的data IO功能）。然后，减去数据均值。然后利用SVD分解，将原有矩阵进行分解。

*M* = *UΣV*T

其中上式中，*Σ*即为奇异矩阵。根据本次作业的要求，我们取前两维的主要特征。然后将所有的点画在图中。另一方面，我们画了一些纵横相交的线，他们共生成25个交叉点。找到理他们最近的数据点，用红色点进行标记即可。

**实验效果：**



作业3. 2D-MOG

**实验原理：**

本次作业介绍了二维的高斯混合模型以及E-M算法

数据生成方面，利用Python自带的线性代数库numpy中的multivariate\_normal函数生成二维的高斯分布。

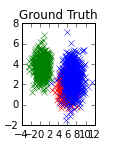
通过E-M算法进行求解。E-M算法分为两个步骤，E步和M步。

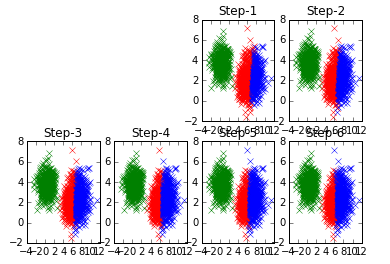
其中E步以当前参数推测因变量的分布，并计算对数似然关于隐变量的期望；

M步，寻找参数最大化期望似然。

这样迭代多步，即可得到最终结果。

**实验结果：**





作业4. Levenberg-Marquardt Algorithm

**实验原理：（From PPT）**

Step1. 给定初始点，*x*0，*μ*0 > 0，*k* = 1

Step2. 计算*gk*和*Gk*。

Step3. 若||*gk*|| < *ε*

Step4. 分解*Gk + μ*kI。若不正定，置*μ*k = 4*μ*k，知道正定为止。

Step5. 解：(*Gk + μ*kI)s = -*gk*，求出*sk*。

Step6. 求f(*x*k+*sk*)，qk(*sk*)。和*rk*=Δ*fk*/Δ*qk*

Step7. 若r < 0.25，置*μ*k+1 = 4*μ*k。若r > 0.75，置*μ*k+1 = *μ*k/2。否则不变；

Step8. 若r < 0，置*x*k+1 = *x*k。否则，*x*k+1 = *x*k + *sk*

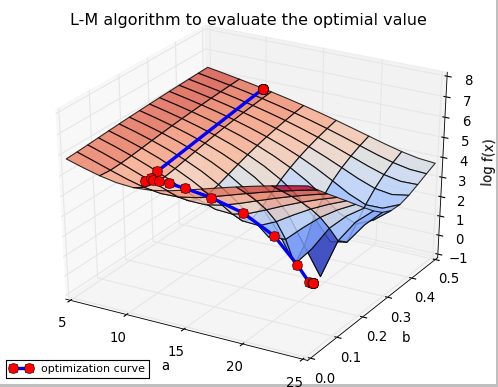
Step9. k += 1;

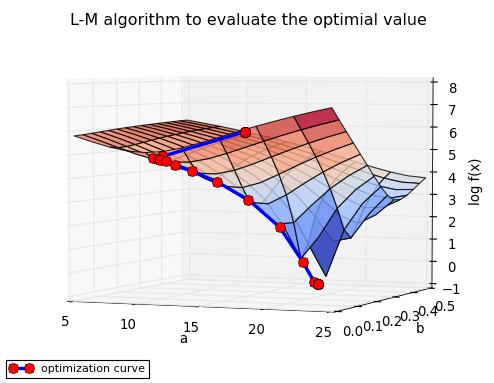
**实验思路：**

因为L-M算法本质上是一个变步长的算法，因此，我们设计一个函数：

并给出10个样本点，通过L-M算法求得最优值。

**实验效果：**





作业5. 2D SVM

**实验原理：**

SVM方法，是通过最优化分割超平面到支持向量的距离，从而求解最优的分割超平面。

SVM不像传统的方法，考虑所有的数据点，而倾向于只考虑离分割超平面最邻近的“支持向量”。在这样的情况下，得到的分类模型是更加鲁棒的。

SVM的模型可以简写为下述内容，我们使用cvxopt这个优化算法包进行求解：

调用Stanford大学Boyd教授的cvxopt最优化库进行求解。最终求解得到。