**最小二乘法曲线拟合参数估计：**

简单起见，这里以一元线性回归为例进行介绍：假设我们获取了一组样本点数据：



利用最小二乘法用多项式曲线拟合这组样本点:

1、设拟合多项式为：



2、样本点到该曲线的距离平方和为：



目标函数为：



上式对参数求偏导有：



则上式等价于：



若X’X的逆矩阵存在：



是一个解析解(即一步到位的解法)

在测试中，我们假设训练样本是在曲线*y*=(*x*2-1)3+(*x*-0.5)2+3sin(2*x*)的基础上对*x*和*y*做随机拉伸处理生成的，我们假设拟合多项式阶数为9（可以用交叉验证来获取），用上述方法拟合得到的结果如图1所示：

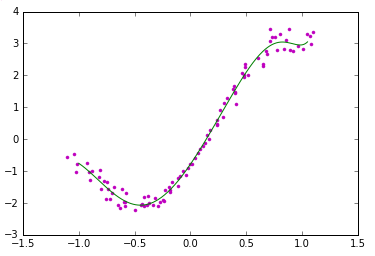


图1 最小二乘法多项式曲线拟合结果

Python代码如下：

**多元线性回归中参数的最大似然估计：**

假设有*m*个训练样本(***x***, *y*)：



假定预测值与样本特征间的函数关系是线性的，即线性回归分析，就在于根据样本***x***和*y*的观察值，去估计函数*h*(***x***)，定义为：



1、对于第个训练样本，假设待估计函数*h*(***x***)与真实值*y*存在误差如下式所示：



假设：误差（即噪声）服从标准正太分布，则有：



2、单个样本下的组样本观测值的概率密度函数为：



3、根据第一篇Logistic Regression文章中最大似然概率的准则可得似然函数如下：



目标函数为：max(*L*(***w***))



4、同第一部分，对上式求偏导数并令其等于0可得：



上式只在逆矩阵存在的时候适用。

在测试中，假设样本生成式为:*y*=0.5*x*+4\*randuniform(0,1)\*sin(2*x*)+3.5，用上述方法做多元线性回归得到的结果下图所示：

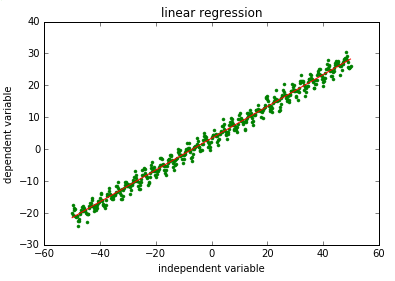


图2 多元线性回归

Python代码如下：