

## 第五讲 课后作业

---

线性回归问题：数据文件 Dataset(5).csv 中包含 9 个训练样本点，组成训练样本集。

- 1) (20)利用第四讲作业实现的线性回归最小二乘法得到训练样本集的最小二乘解，并且利用得到的线性回归模型生成 5 个测试样本，组成测试样本集，使用两种颜色绘出训练样本集和测试样本集中样本的散布图。假设噪声模型为  $\mathcal{N}(0,5)$ 。
- 2) (30)编程实现岭回归算法，求解出训练样本集的岭回归模型，给出平均训练误差和平均测试误差。
- 3) (20)比较岭回归模型与线性回归模型（最小二乘解）的平均训练误差和平均测试误差，给出结论及理由。
- 4) (30)如果使用多项式回归模型来拟合训练样本集，应该选择多少阶的多项式？说明你给出此答案的理由。