

## 任务二：线性回归 (程序请写清楚注释，注意规范)

请将源程序和结果的 pdf 版到 github 上，每个人上传一个文件夹到我的 github 文件夹 task\_2 下，命名格式：ex1\_你的编号 eg: ex1\_01

(1) 前三个题选一个：

你的编号除以三，余数+1 为你所要做的题目 eg: 01 做第 2 道题

(2) 后两个题选一个：

奇数选第一个，偶数选第二个

**## 注意：**以下所有数据集都随机打乱，（将顺序打乱），并将 **80%分为训练集，20%分为测试集**，训练集用来训练模型，测试集用来检验模型

检验结果表示为错误率的形式：

计算公式如下：

$$\frac{1}{n} \sum_{i=1}^n (h(x^*) - y^*)^2$$

（其中 n 表示测试集的样本数， $x^*$ ， $y^*$ 表示测试集的样本

在 pdf 中请包括（1）绘制每次迭代损失函数 J 的变化曲线图（2）模型的参数估计值

（3）测试集上的错误率 （4）\*可选：可视化结果（仅针对二维数据的可视化）

(如下的  $w_i$  为参数，类似于讲的  $\theta$ .)

➤ （1）多项式回归

$$h = w_1 * x + w_2 * x.^2 + w_3 * x.^3 + w_4 * x.^4 + w_5 * x.^5 + w_6$$

多项式最大次方设置为 5

➤ （2）对 x 使用基函数先进行变换（原来是由 sin 函数产生）基变换

$$h = \sin(w_1 * x) + \sin(w_2 * x) + w_3$$

➤ （3）多维线性回归

$$h = w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + w_4$$

➤ （4）二维数据的可视化（产生二维数据高斯分布）

要求：设计函数使之可以输入均值，方差来指定特定的高维高斯分布

(1) 三维图（坐标轴分别为 x,y 和概率密度 p）

(2) 等高线图 (hint: contour, contourf)

hint：使用 mvnpdf 可以产生如：均值为[0 0],方差为[2 0;0 4]的高斯分布

(3) 画出均值为[0 0],方差为[2 0;0 4]的上述图片，并附在 pdf 中

(4)\* 其他可能的可视化图

➤ （5）预测汽车的价格（见真实数据集）

通过汽车的 25 种特征（不一定全都要使用），预测汽车的价格（price）

尽量想办法尝试提高准确度（如：怎么处理缺失值，怎么处理非连续的变量）