

# 代码使用说明及模型的介绍

## 1. 代码文件夹及其代码文件的介绍

代码文件夹内容介绍：

代码文件夹中共有 1 个文件夹，分别是 data。

data 文件夹下存放的是训练集数据 data.xlsx 和测试集数据 100\_test.xlsx（是最新的测试集数据）。其中 data.xlsx 的 sheet 名称为 train\_400，即 400 个训练数据，100\_test.xlsx 的 sheet 名称为 valid\_data，即 100 个测试数据。

代码介绍：

由于模型训练较快，因此将模型的训练和测试统一写成了 ACO\_model.m 脚本文件。其中模型对应的函数脚本文件为 trainAnts.m, fmincg.m, linearRegCostFunction.m。

## 2. 运行代码方法

- 直接打开 matlab 软件运行 ACO\_model.m 脚本文件。
- 程序就会根据输入的模型进行训练，最后输出验证集的预测准确率。

## 3. 验证集准确率介绍

- 将验证集准确率定义为模型输出和验证集输出值偏差在 5%范围内的数据占比。
- 模型输出和验证集输出值偏差定义为二者的绝对值之差与标准输出的占比。
- 比如模型输出为 90，标准输出为 96，则偏差为 $(96-90)/96 = 0.0625$ ，即 6.25%。
- 而验证集中数据有 30 个满足以上偏差范围，则准确率为  $30/40=75\%$ 。

## 4. 模型介绍

- 模型使用线性映射（非线性化）的最小二乘法的回归模型，同时训练中使用蚁群智能算法。首先将数据按维度进行线性映射，使其适应非线性模型，最后使用蚁群智能算法和梯度下降法进行拟合。蚁群智能算法使用在最优化问题中，因此只适用于训练中，在最小化损失函数(最小二乘法是最小化均方误差)条件下，使用该算法寻找全局最优解，即蚁群智能算法寻求最能拟合训练数据的模型，其中添加正则化项用参数  $\lambda$  进行控制和探索，参数  $\lambda$  的大小意味着模型对于正则化项的重视程度。该正则化项防止过拟合，而使模型不具有很好的泛化性。

## 5. 模型结果分析

- 经过多次训练后固定参数权重得出，ACO 模型的训练集的准确率为 80%，测试集的准确率为 87%。
- 蚁群智能算法旨在寻求最能拟合训练数据的模型，但容易过拟合，泛化性能不是很好。

## 6. 模型训练参数

- + 信息遗留因子  $\alpha$  : 0.5
- + 蚁群蚂蚁个数 m: 50
- + 信息素增强系数 Q: 1500.0
- + 信息素挥发系数  $\rho$  : 0.7

+ 训练最大回合数: 50