



成都大学

CHENGDU UNIVERSITY

《人工智能导论》
实验指导书

杨耀如

成都大学 电子信息与电气工程学院

实验二 逻辑回归

一、实验目的

掌握用 MATLAB 编写代码实现逻辑回归的方法，加深对逻辑回归的原理，从建模、训练到使用的过程的理解。

二、实验内容

使用 MATLAB，完成代码模版，实现逻辑回归涉及到的各个函数功能，并绘制逻辑回归分类图。

三、实验设备

- 1、计算机
- 2、MATLAB
- 3、Lab2 MATLAB 代码模版

四、实验原理

1、代码模版

本实验使用的 MATLAB 代码模版共包含以下文件：

- (1) lab2.m 本文件会指引你完成本实验
- (2) plotrawdata.m 绘制 2-d 分类数据图像的函数
- (3) sgmd.m sigmoid 函数
- (4) objllfunction.m 建立逻辑回归模型的似然目标函数
- (5) forecast.m 使用逻辑回归分类器进行预测的函数
- (6) plotclassifier.m 绘制决策边界
- (7) lab2data.txt 数据集

其中 (2) - (6) 中的业务逻辑部分需要完成

2、逻辑回归

在本实验中，你将实现一个逻辑回归分类器，并将之运用于预测。

假如你是一个考官，需要决定每位申请者是否能够成功通过测试的概论，依据为申请者两门测试的成绩。你可以查询的依据为以往申请者的历史数据，这些数据可以用来做为训练集来训练逻辑回归模型。对于每一个训练数据，由两门测试的成绩和是否通过的决策这三个属性组成。lab2.m 作为该使用的框架，将引导你完成本实验。

3、数据可视化

开始建立任何模型前，如果可能的话，将数据可视化，对数据得到直观的印象总是好的。在 lab2.m 中的代码会通过执行 plotrawdata 函数来实现这一功能。

请完成 plotrawdata.m，使得该函数可以支持做出如图 1 所示的图像，其中两个坐标轴分别为两个测试的分数，正类样本和负类样本使用不同的记号。

4、sigmoid 函数

逻辑回归通过 sigmoid 函数实现了将特征与权重的线形组合映射为概论值的功能：

$$h_{\theta}(x) = g(\theta^T x)$$

其中, g 为 sigmoid 函数, sigmoid 函数本身定义为:

$$g(z) = \frac{1}{1 + e^{-z}}$$

完成 `sgmd.m`, 实现 sigmoid 函数。

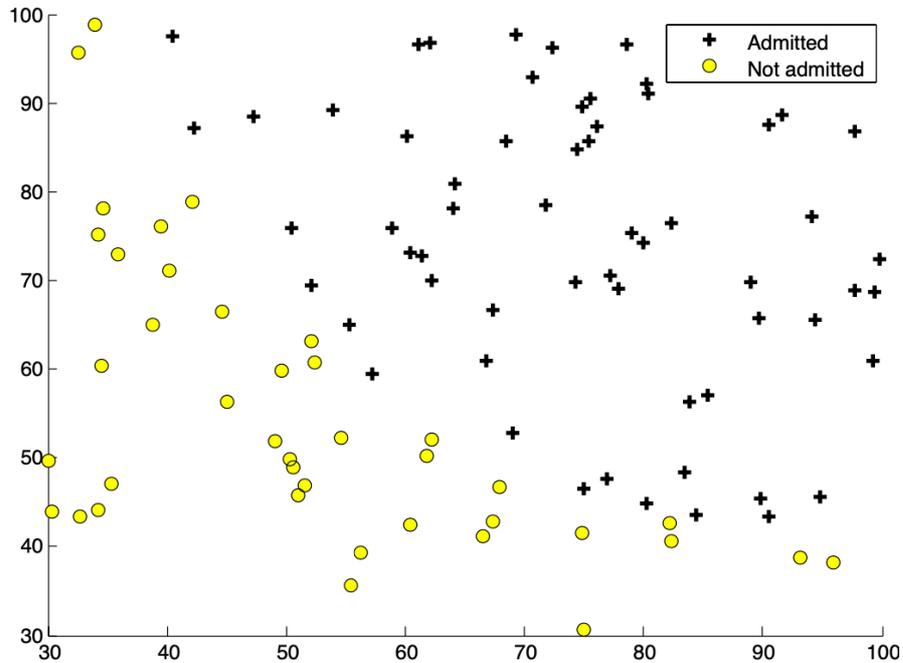


图 1 数据可视化

5、似然目标函数和梯度

完成 `objllfunction.m`, 该函数计算逻辑回归的似然目标函数和梯度, 将计算结果返回。我们使用如下函数为逻辑回归的目标函数:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(h_{\theta}(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

其梯度为与 θ 等长的一个向量:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

完成 `objllfunction.m` 后, 你可以运行 `lab2.m`, `lab2.m` 会调用你的 `objllfunction.m` 函数, 使用初始化后的参数 θ 计算, 你可以观察命令窗口打印的代价值, 为 .693。

6、使用 `fminunc` 学习参数

在 `lab2.m` 中, 使用正确的参数去调用 `fminunc` 的代码已经写好。

具体地说, 在给定数据集 (给定 \mathbf{X} 和 \mathbf{y}) 的情况下, 你需要使用 `fminunc` 去找到当前逻辑回归模型最优的参数 θ , 你需要给 `fminunc` 传递如下参数:

- (1) 待优化的参数的初始值;
- (2) 一个可以在给定数据集和待优化参数时计算逻辑回归的似然目标函数和梯度的函数。

在这个代码片段中（见图 2），我们首先定义 *fminunc* 需要用到的选项：我们将 *GradObj* 选项设置为 “on”，这相当于告诉函数我们的函数既返回似然目标函数值和梯度值。这使得 *fminunc* 在最小化目标函数时使用梯度为依据；我们设置 *MaxIter* 选项为 400，是 *fminunc* 在终止前运行最多 400 次。

为了告诉 *fminunc* 具体需要优化哪个函数哪个参数，我们使用了特定的形式 *@(t)(objllfunction(t,X,y))*。这样会隐式地生成一个带参数 *t* 的函数去调用你的 *objllfunction* 函数。

如果你正确地完成了 *objllfunction.m*，*fminunc* 会收敛于正确的优化参数，并返回似然目标函数的最终值和参数 θ 。注意在使用 *fminunc* 时，你不用自己书写任何的循环，也不用尝试不同的学习率参数，这些都由 *fminunc* 自动完成了，你要做的只是提供给 *fminunc* 一个可以计算目标函数值和梯度的函数。

一旦 *fminunc* 执行完成，*lab2.m* 会使用最优的参数 θ 调用 *objllfunction* 函数。你会观察到最终的目标函数值约为 .203。

```
% Set options for fminunc
options = optimset('GradObj', 'on', 'MaxIter', 400);

% Run fminunc to obtain the optimal theta
% This function will return theta and the cost
[theta, cost] = ...
    fminunc(@(t)(costFunction(t, X, y)), initial_theta, options);
```

图 2 使用 *fminunc* 的代码片段

最终的参数 θ 会被用于绘制决策边界，完成 *plotclassifier.m* 文件，并继续运行 *lab2.m*，结果大约如图 3 所示。

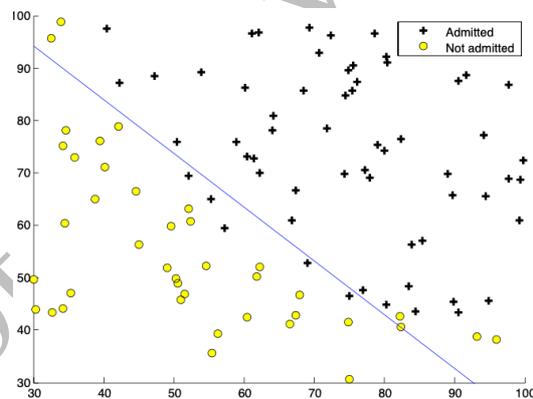


图 3 分类决策边界

7、评估逻辑回归模型

在学得参数后，该逻辑回归模型可以用于预测申请人是否会通过测试。对于一个持有第一门测试 45 分，第二门测试 85 分的成绩的申请人，通过本逻辑回归模型预测，你会在命令行窗口观察到通过概率约为 .776。

完成 *forecast.m* 实现评估功能。

五、实验要求

- 1、编写 MATLAB 代码，完成四.1 中提及的函数（2）-（6）；
- 2、绘制原始数据分布图形和待油分类边界的图形；
- 3、分析实验结果。

Intro to AI by Y YANG