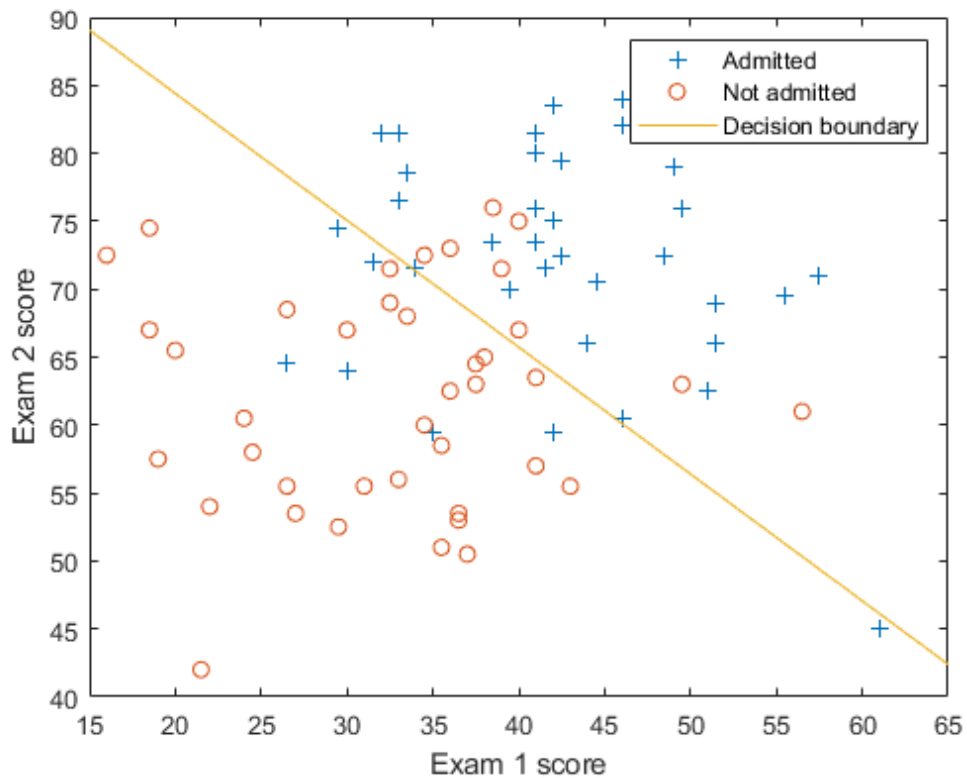


山东大学计算机科学与技术学院

机器学习与模式识别课程实验报告

学号：201600181073	姓名：唐超	班级：智能 16
实验题目：Logistic Regression and Newton's Method		
实验学时：2	实验日期：10.22	
实验目的： 用逻辑回归方法实现一个高校基于成绩数据是否给予录取的分类问题。		
硬件环境： DELL 台式机		
软件环境： MATLAB R2018b		
实验步骤与内容： <ol style="list-style-type: none"> 1. 下载、解压并加载数据至 MATLAB，成绩数据命名为 x，是否录取的标签命名为 y； 2. 在矩阵 x 的第一列前插入 80*1 的全部元素为 1 的向量； 3. 将 sigmoid 函数作为内置函数 g： $g = \text{inline}('1.0./(1.0+\exp(-z))');$ 4. 采用下列公式多次迭代： $\begin{cases} grad = ((g(\theta^T \cdot x^T) - y^T) \cdot x)^T \\ H = x^T \cdot \text{diag}(g(\theta^T \cdot x^T) \cdot g(-\theta^T \cdot x^T)) \cdot x \\ \theta^{(t+1)} = \theta^{(t)} - H^{-1} \cdot grad \end{cases}$ <p>经过 6 次迭代，θ 已经收敛，$\theta = \begin{pmatrix} -16.3787 \\ 0.1483 \\ 0.1589 \end{pmatrix}$</p> 5. 设横坐标组成的向量 $m = 15:65$，通过 $\theta^T x = 0$，解出相应的纵坐标向量 n，画出两种类别的分界线，结果如下图所示： 		



6. 对学生 (1,20,80) 是否被录取进行预测，经计算，其不被录取的概率为

$$p = 1 - g(\theta' \cdot (1, 20, 80)') = 0.6680$$

结论分析与体会：

通过这次实验加深了对逻辑回归的理解。

附录：程序源代码

```
x = load('ex4x.dat');
y = load('ex4y.dat');
x = [ones(80, 1), x];

pos = find(y == 1);
neg = find(y == 0);

plot(x(pos, 2), x(pos, 3), 'r+');
hold on
```

```

plot(x(neg,2),x(neg,3),'o');

g = inline('1.0./(1.0+exp(-z))');

theta = zeros(3,1);

for i = 1:6
    grad = ((g(theta'*x')- y')*x)'; %维度 3*1
    H = x'*diag(g(theta'*x')).*g(-theta'*x')*x;
    theta = theta - H\grad;
end

%画出分界线
m = 15:65;
n = zeros(1,51);
for i = 1:51
    n(i) = -(theta(1)+theta(2)*m(i))/theta(3);
end

plot(m,n);
xlabel('Exam 1 score');
ylabel('Exam 2 score');
legend('Admitted','Not admitted','Decision boundary');

p = [1;20;80];
probability_p = 1 - g(theta*p) %not admitted

```

Published with MATLAB® R2018b