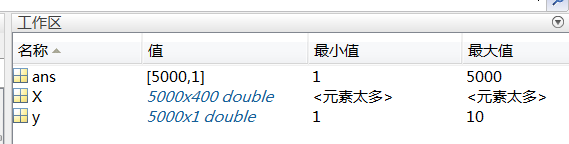
[stanford coursera 机器学习编程作业 exercise 3（逻辑回归实现多分类问题）](https://www.cnblogs.com/hapjin/p/6085278.html)

本作业使用逻辑回归(logistic regression)和神经网络(neural networks)识别手写的阿拉伯数字(0-9)

关于逻辑回归的一个编程练习，可参考：[Stanford coursera Andrew Ng 机器学习课程编程作业（Exercise 2）及总结](http://www.cnblogs.com/hapjin/p/6078530.html)

下面使用逻辑回归实现多分类问题：识别手写的阿拉伯数字(0-9)，使用神经网络实现：识别手写的阿拉伯数字(0-9)，请参考：[神经网络实现](http://www.cnblogs.com/hapjin/p/6085489.html)

数据加载到Matlab中的格式如下：

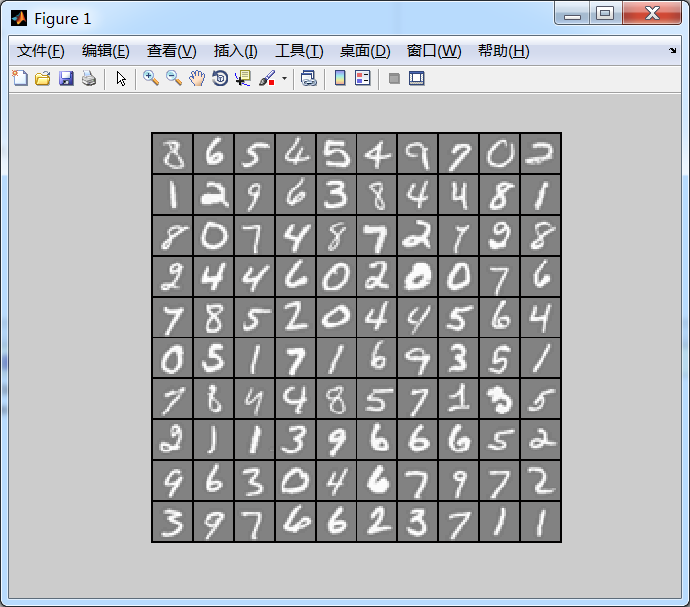


一共有5000个训练样本，每个训练样本是400维的列向量（20X20像素的 grayscale image），用矩阵 X 保存。样本的结果(label of training set)保存在向量 y 中，y 是一个5000行1列的列向量。

比如 y = (1,2,3,4,5,6,7,8,9,10......)T，注意，由于Matlab下标是从1开始的，故用 10 表示数字 0

**①样本数据的可视化**

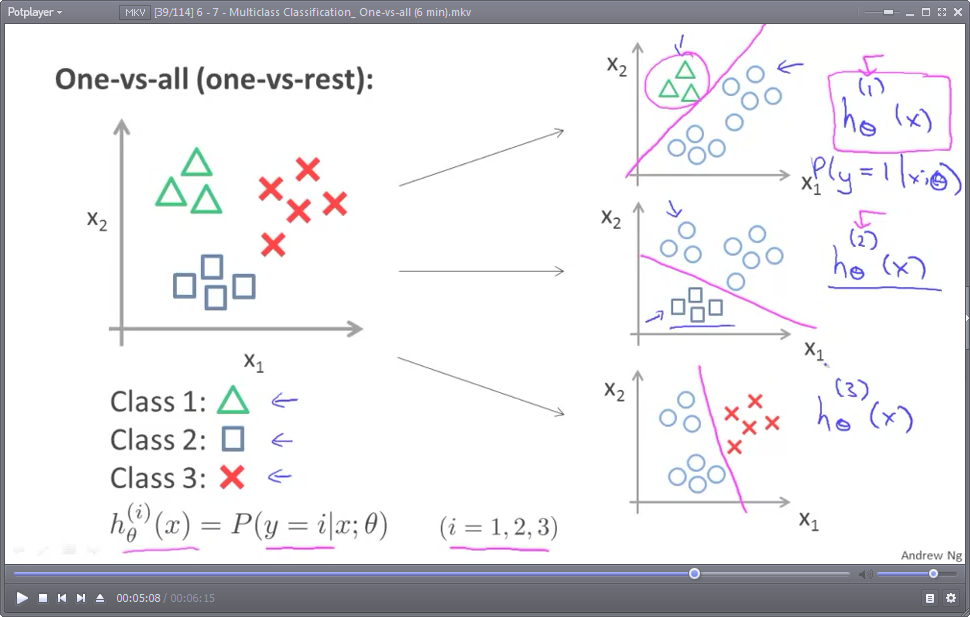
随机选择100个样本数据，使用Matlab可视化的结果如下：



**②使用逻辑回归来实现多分类问题(one-vs-all)**

所谓多分类问题，是指分类的结果为三类以上。比如，预测明天的天气结果为三类：晴(用y==1表示)、阴(用y==2表示)、雨(用y==3表示)

分类的思想，其实与逻辑回归分类(默认是指二分类，binary classification)很相似，**对“晴天”进行分类时，将另外两类(阴天和下雨)视为一类**：(非晴天)，这样，就把一个多分类问题转化成了二分类问题。示意图如下：（图中的圆圈 表示：不属于某一类的 所有其他类）



对于N分类问题(N>=3)，就需要N个假设函数(预测模型)，也即需要N组模型参数θ（θ一般是一个向量）

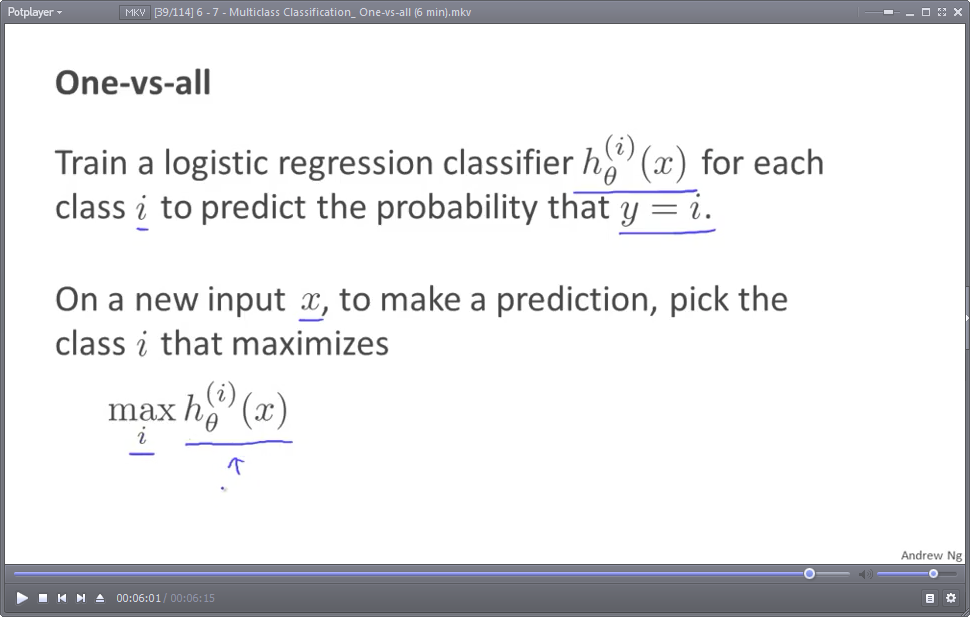
然后，对于每个样本实例，依次使用每个模型预测输出，选取输出值最大的那组模型所对应的预测结果作为最终结果。

因为模型的输出值，在sigmoid函数作用下，其实是一个概率值。，注意：hθ(1)(x)，hθ(2)(x)，hθ(3)(x)三组 模型参数θ 一般是不同的。比如：

hθ(1)(x)，输出 预测为晴天(y==1)的概率

hθ(2)(x)，输出 预测为阴天(y==2)的概率

hθ(3)(x)，输出 预测为雨天(y==3)的概率



**③Matlab代码实现**

对于上面的识别阿拉伯数字的问题，一共需要训练出10个逻辑回归模型，每个逻辑回归模型对应着识别其中一个数字。

我们一共有5000个样本，样本的预测结果值就是：y=(1,2,3,4,5,6,7,8,9,10)，其中 10 代表 数字0

我们使用Matlab fmincg库函数 来求解使得代价函数取最小值的 模型参数θ

[复制代码](javascript:void(0);)

function [all\_theta] = oneVsAll(X, y, num\_labels, lambda)

%ONEVSALL trains multiple logistic regression classifiers and returns all

%the classifiers in a matrix all\_theta, where the i-th row of all\_theta

%corresponds to the classifier for label i

% [all\_theta] = ONEVSALL(X, y, num\_labels, lambda) trains num\_labels

% logisitc regression classifiers and returns each of these classifiers

% in a matrix all\_theta, where the i-th row of all\_theta corresponds

% to the classifier for label i

% Some useful variables

m = size(X, 1);% num of samples

n = size(X, 2);% num of features

% You need to return the following variables correctly

all\_theta = zeros(num\_labels, n + 1);

% Add ones to the X data matrix

X = [ones(m, 1) X];

% ====================== YOUR CODE HERE ======================

% Instructions: You should complete the following code to train num\_labels

% logistic regression classifiers with regularization

% parameter lambda.

%

% Hint: theta(:) will return a column vector.

%

% Hint: You can use y == c to obtain a vector of 1's and 0's that tell use

% whether the ground truth is true/false for this class.

%

% Note: For this assignment, we recommend using fmincg to optimize the cost

% function. It is okay to use a for-loop (for c = 1:num\_labels) to

% loop over the different classes.

%

% fmincg works similarly to fminunc, but is more efficient when we

% are dealing with large number of parameters.

%

% Example Code for fmincg:

%

% % Set Initial theta

% initial\_theta = zeros(n + 1, 1);

%

% % Set options for fminunc

% options = optimset('GradObj', 'on', 'MaxIter', 50);

%

% % Run fmincg to obtain the optimal theta

% % This function will return theta and the cost

% [theta] = ...

% fmincg (@(t)(lrCostFunction(t, X, (y == c), lambda)), ...

% initial\_theta, options);

%

initial\_theta = zeros(n + 1, 1);

options = optimset('GradObj','on','MaxIter',50);

for c = 1:num\_labels %num\_labels 为逻辑回归训练器的个数，num of logistic regression classifiers

all\_theta(c, :) = fmincg(@(t)(lrCostFunction(t, X, (y == c),lambda)), initial\_theta,options );

end

% =========================================================================

end

[复制代码](javascript:void(0);)

lrCostFunction，完全可参考：http://www.cnblogs.com/hapjin/p/6078530.html 里面的 正则化的逻辑回归模型实现costFunctionReg.m文件

下面来解释一下 for循环：

num\_labels 为分类器个数，共10个，每个分类器(模型)用来识别10个数字中的某一个。

我们一共有5000个样本，每个样本有400中特征变量，因此：模型参数θ 向量有401个元素。

initial\_theta = zeros(n + 1, 1); % 模型参数θ的初始值(n == 400)

all\_theta是一个10\*401的矩阵，每一行存储着一个分类器(模型)的模型参数θ 向量，执行上面for循环，就调用fmincg库函数 求出了 所有模型的参数θ 向量了。

求出了每个模型的参数向量θ，就可以用 训练好的模型来识别数字了。对于一个给定的数字输入(400个 feature variables) input instance，每个模型的假设函数hθ(i)(x) 输出一个值(i = 1,2,...10)。取这10个值中最大值那个值，作为最终的识别结果。比如g(hθ(8)(x))==0.96 比其它所有的 g(hθ(i)(x)) (i = 1,2,...10,但 i 不等于8) 都大，则识别的结果为 数字 8

[复制代码](javascript:void(0);)

function p = predictOneVsAll(all\_theta, X)

%PREDICT Predict the label for a trained one-vs-all classifier. The labels

%are in the range 1..K, where K = size(all\_theta, 1).

% p = PREDICTONEVSALL(all\_theta, X) will return a vector of predictions

% for each example in the matrix X. Note that X contains the examples in

% rows. all\_theta is a matrix where the i-th row is a trained logistic

% regression theta vector for the i-th class. You should set p to a vector

% of values from 1..K (e.g., p = [1; 3; 1; 2] predicts classes 1, 3, 1, 2

% for 4 examples)

m = size(X, 1);

num\_labels = size(all\_theta, 1);

% You need to return the following variables correctly

p = zeros(size(X, 1), 1);

% Add ones to the X data matrix

X = [ones(m, 1) X];

% ====================== YOUR CODE HERE ======================

% Instructions: Complete the following code to make predictions using

% your learned logistic regression parameters (one-vs-all).

% You should set p to a vector of (from 1 to

% num\_labels).

%

% Hint: This code can be done all vectorized using the max function.

% In particular, the max function can also return the index of the

% max element, for more information see 'help max'. If your examples

% are in rows, then, you can use max(A, [], 2) to obtain the max

% for each row.

%

[~,p] = max( X \* all\_theta',[],2); % 求矩阵(X\*all\_theta')每行的最大值，p 记录矩阵每行的最大值的索引

% =========================================================================

end

[复制代码](javascript:void(0);)