Problem 1: Logistic regression

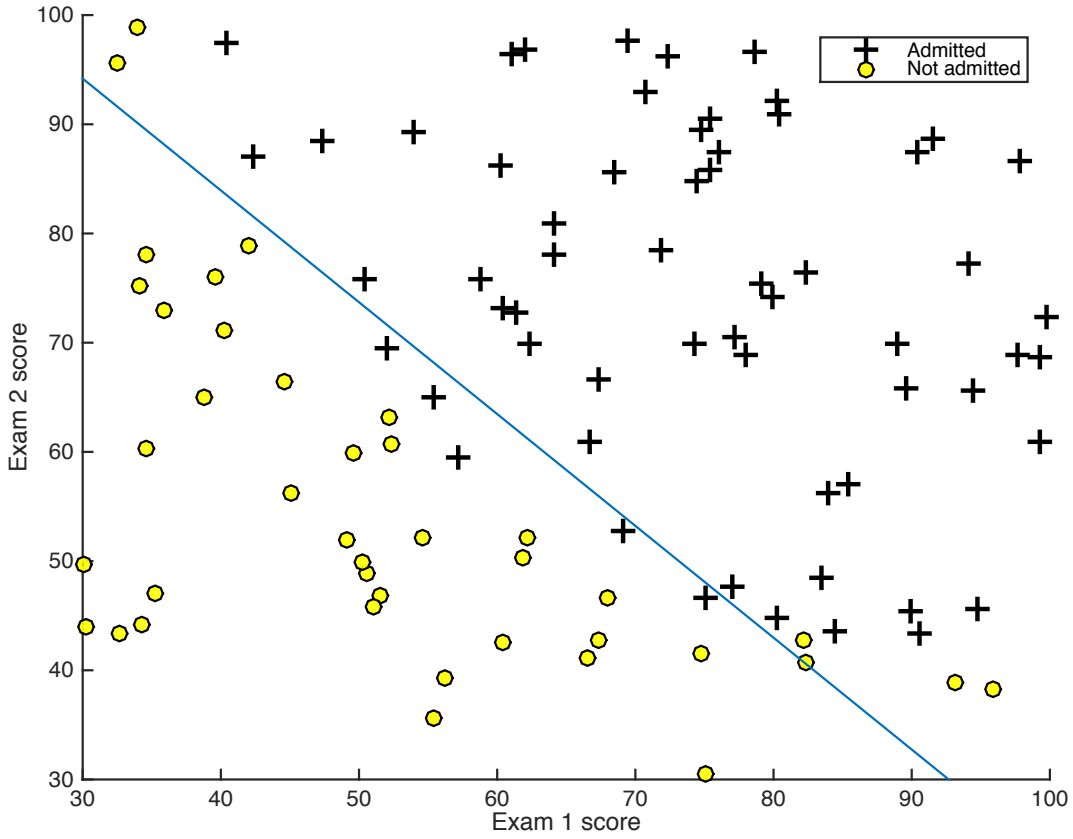
Implementing logistic regression : the sigmoid function

Cost function and gradient of logistic regression

Gradient at initial theta (zeros):
-0.100000
-12.009217
-11.262842

Learning parameters using fminunc

Cost at theta found by fminunc: 0.203506
theta:
-24.932780
0.204406
0.199616



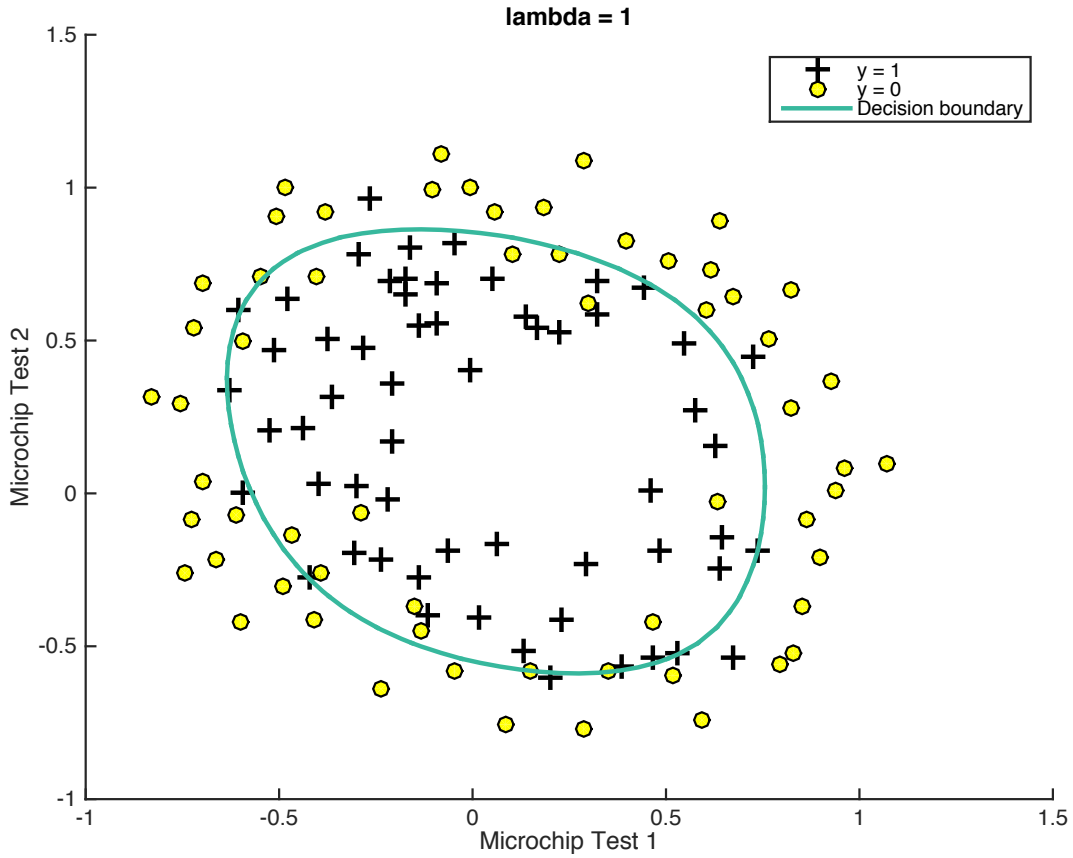
Evaluating logistic regression
For a student with scores 45 and 85, we predict an admission probability of 0.774321

Train Accuracy: 89.000000

Problem 2: Regularized logistic regression

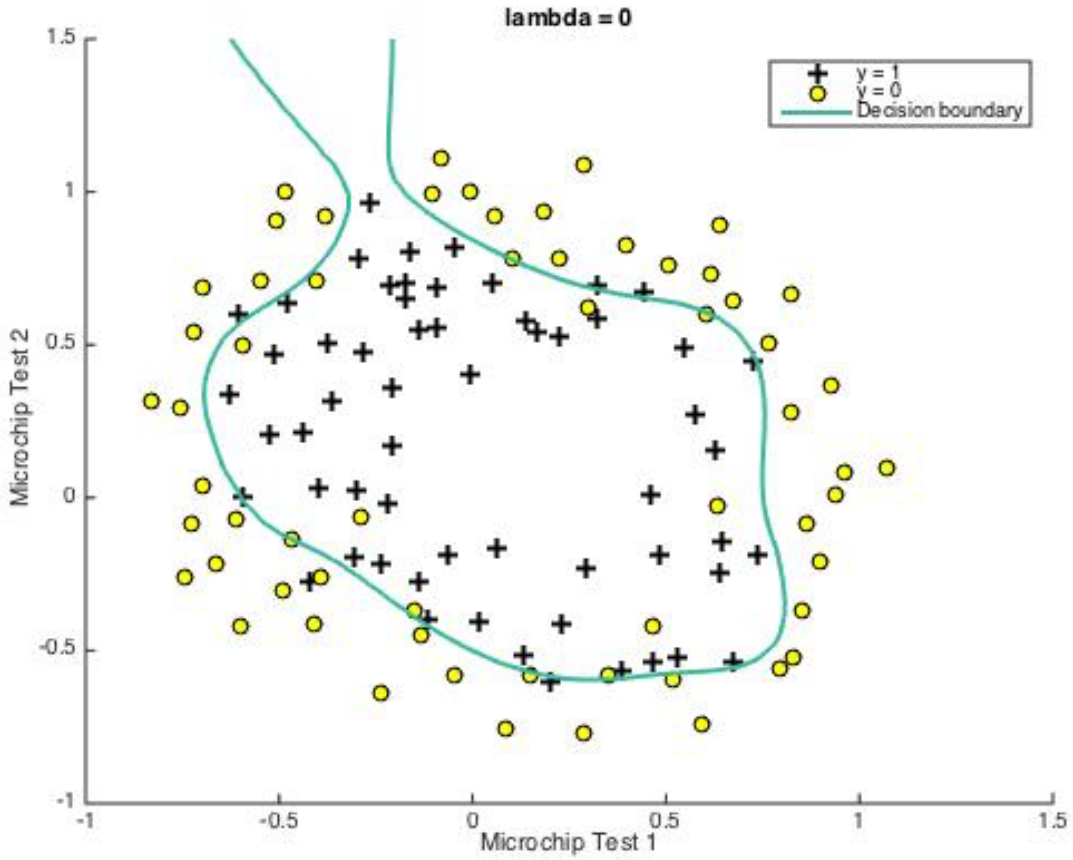
Cost function and gradient
Train Accuracy: 83.050847

Plotting the decision boundary

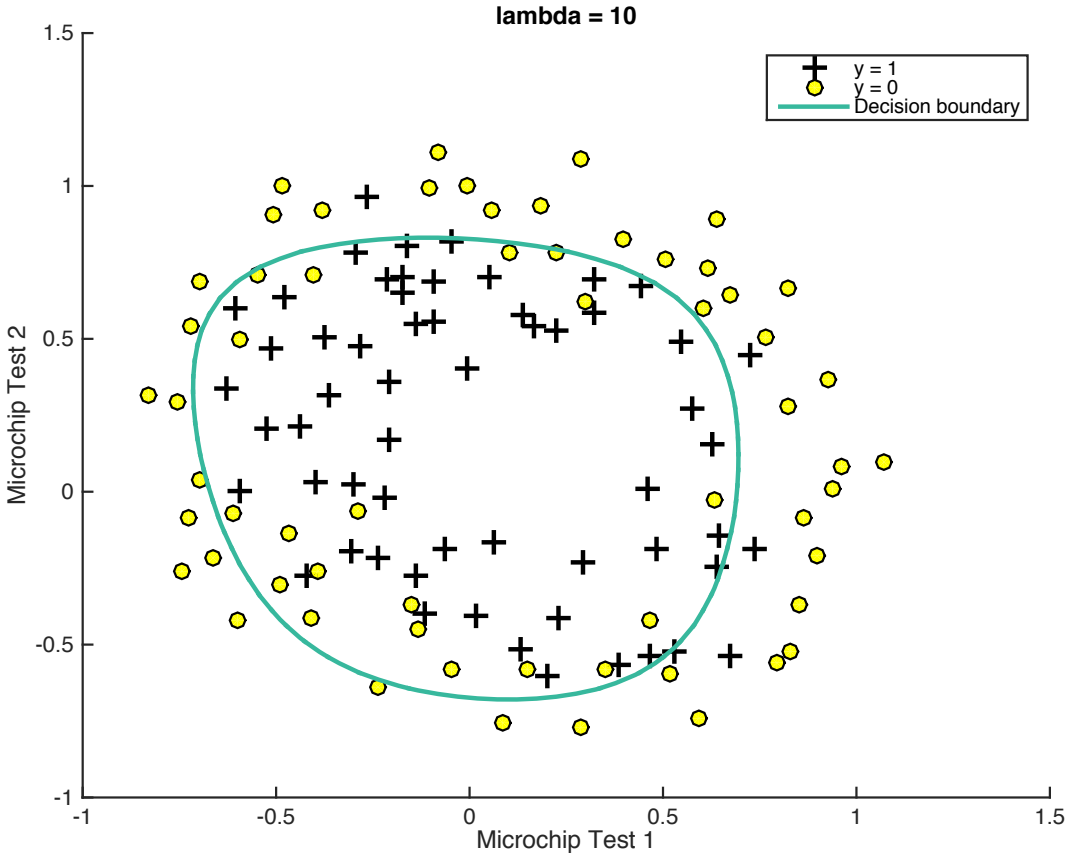


Varying lambda

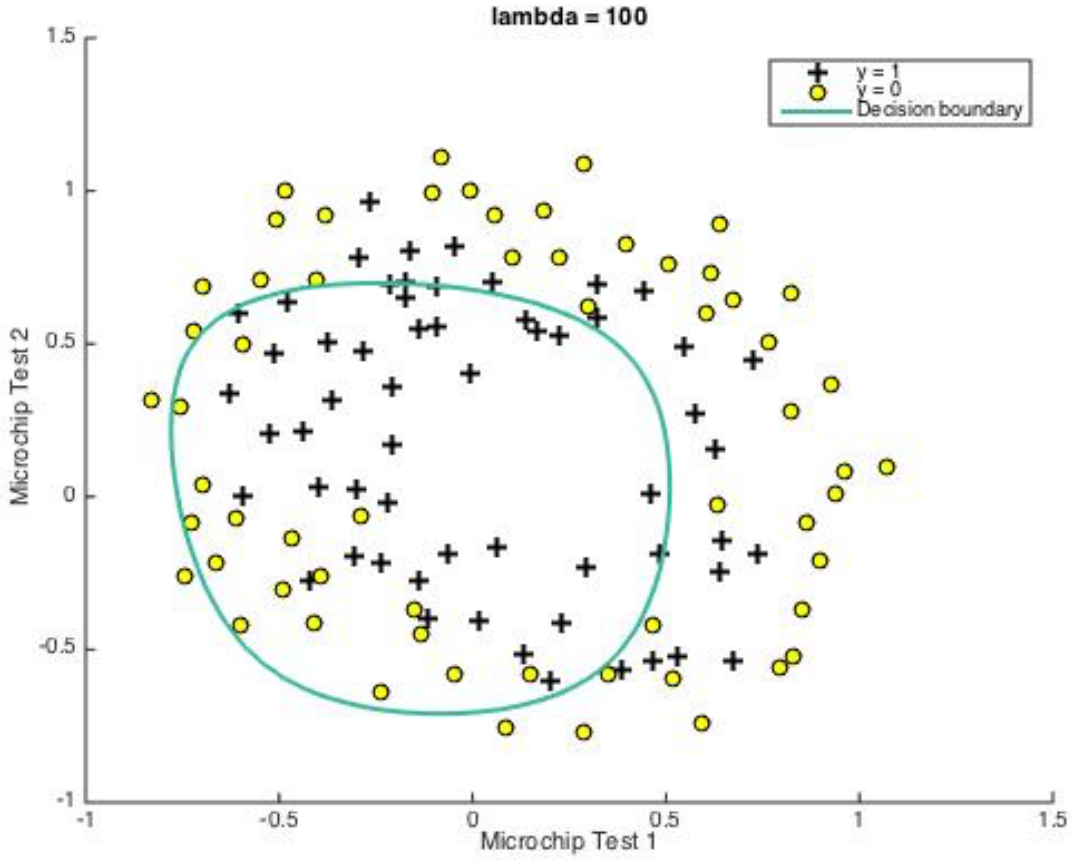
lambda = 0, overfitting, train accuracy: 88.135593



lambda = 10, overfitting, train accuracy: 79.661017



lambda = 100, underfitting, train accuracy: 66.949153



Problem 3: Logistic regression for spam classification

Fitting regularized logistic regression models

binarizeFeatures
Train Accuracy with lambda = 3.500000: 93.572594
Test Accuracy with lambda = 3.500000: 92.708333

stdFeatures
Train Accuracy with lambda = 4.500000: 91.973899
Test Accuracy with lambda = 4.500000: 91.341146

logTransformFeatures
Train Accuracy with lambda = 0.500000: 94.747145
Test Accuracy with lambda = 0.500000: 94.205729