

Machine Learning – WS'12

Exercise-4

Prof. Dr. Dr. Lars Schmidt-Thieme, Umer Khan
Information Systems and Machine Learning Lab (ISMLL),
University of Hildesheim

Logistic Regression:

Problem-1:

Given the following data:

| y | x | y | x |
|---|------|---|------|
| 0 | 9.5 | 1 | 11.1 |
| 0 | 9.6 | 1 | 11.1 |
| 0 | 9.7 | 1 | 11.1 |
| 0 | 9.8 | 1 | 11.5 |
| 0 | 9.9 | 1 | 11.8 |
| 0 | 10.5 | 1 | 11.9 |
| 0 | 11.0 | 1 | 12.1 |
| 0 | 11.2 | 1 | 12.2 |
| 0 | 11.5 | 1 | 12.5 |
| 0 | 11.7 | 1 | 12.6 |
| 0 | 12.1 | 1 | 12.6 |

Suppose the following problem: A doctor wants to find out whether a particular antibiotic has an effect on the incidence of infection in women who have a caesarean section. He starts from a simple linear regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$y = 1$ means that an infection occurs in the next 2 weeks, $y = 0$, that no infection occurs. x encoding the administered amount of the antibiotic. When the target variable is a binomial random variable with the following discrete distribution:

$$p(y) = 1 - p \text{ wenn } y = 0$$

$$p(y) = p \text{ wenn } y = 1$$

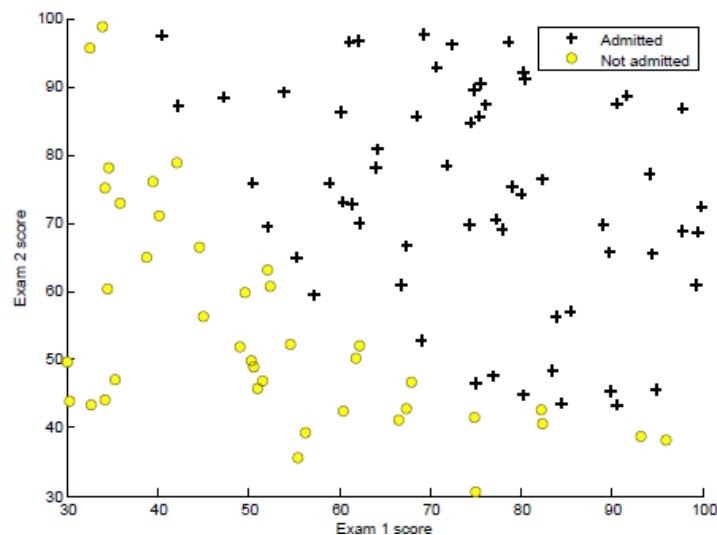
a) Why should you not use linear regression on data with binary outcome variable? Consider the error 'e' and the variance σ^2 . What value ϵ takes for $y = 1$ and $y = 0$? What value does σ^2 takes?]

b) We assume that the variance for data x can be varying in different cases. What method can be applied to solve the problem?

c) Suppose we use logistic regression. Describe the behavior of the logistic function with respect to β

Problem 2:

Build a logistic regression model to predict whether a student gets admitted into a university based on 'Exam 1 Score' and 'Exam 2 Score'. You have to estimate the probability of a student to get admitted. Implement 'plotData.m' so that the training data in 'ex2data1.txt' can be visualized as in the following figure:



Implement a Sigmoid Function:

Recall that logistic regression hypothesis is defined as:

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

Where function 'g' is a sigmoid function defined as:

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

Implement 'sigmoid.m'. Test sigmoid(x) for few values. For very large positive values of x, sigmoid should be close to 1, and for large negatives it should be close to 0. At x=0, it should be exactly 0.5.

For a matrix, your function should perform sigmoid function on every element.

Cost Function and Gradient:

Implement 'costFunction.m' to return cost and gradient of Logistic Regression, which takes the following form:

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

And the gradient of the cost is a vector ' Θ ' (theta) where the jth element (for j= 1,2,...,n) is defined as:

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

Test your costFunction through 'ex2.m' using the initial parameters of ' Θ '

Learning Parameters through 'fminunc':

'fminunc' is an optimization solver built-in Octave, which finds the minimum of an unconstrained function. Logistic regression is unconstrained since ' Θ ' can take any real value. It takes initial parameters ' Θ ', the costFunction which returns cost and gradient for your data (X,y). 'ex2.m' has already the code to call 'fminunc' (Also explained in tutorial). You will notice that you do not have to specify any loop and learning rate. 'fminunc' will do this automatically. Finally it returns optimized 'theta' and value of cost function. 'ex2.m' will then call 'plotDataBoundary.m' to plot a decision boundary on training data using ' Θ s'. The code for 'plotDataBoundary.m' is already given, but you should try to understand how such boundaries can be plotted in Octave.

Evaluating Logistic Regression:

After learning the parameters, you can use the model to predict whether a particular student will be admitted. For a student with an Exam 1 score of 45 and an Exam 2 score of 85, you should expect to see an admission probability of 0.774. Complete the code in 'predict.m' which produces '1' or '0' and takes data set and thetas as parameters. Using 'predict.m', 'ex2.m' will report training accuracy of classifier based on percentage of examples it classified correctly.