Machine Learning for Computer Vision (EE5177)
Programming Assignment 4 : Logistic Regression and SVMs
Problem #1

Akshit Kumar
EE14B127

29th March 2017

Contents

1 Introduction

1.1 Goal

The goal of the question is to generate synthetic data of 2 classes of univariate normal distributions with the following parameters:

- Class 0 : Mean = -2 , Standard Deviation = 1.5
- Class 1 : Mean = 2 , Standard Deviation = 1.5

The goal is to find a separating hyperplane for the generated data points for which we use Logistic Regression. The data is not linearly separable which is also evident from the plot.

1.2 Approach

To find the separating hyperplane, during the training phase, we try to minimise the negative log likelihood function. For minimising the log likelihood function and getting the optimal value of ϕ, we use the *fminunc* MATLAB function which takes in the cost function, the gradient of the cost function and the hessian of the cost function.

2 Data Generation

The data is synthesized by randomly sampling from univariate Gaussian distribtuions:

- Class 0 : $\mathcal{N}(-2, 1.5)$
- Class 1 : $\mathcal{N}(2, 1.5)$

The data is split in the ratio of 80/20 for training and testing respectively.

3 Training Phase

3.1 Goal

The goal of the training phase is to come up with optimal value of ϕ which minimises the negative of the log likelihood function to learn the separating hyperplane.

3.2 Approach

The training phase for logistic regression model tries to minimise the negative log likelihood function using the *fminunc* MATLAB function which takes into account the cost function, gradient of the cost function and hessian of the cost function. The required functions which are passed into the *fminunc* function are :

- Cost Function : $L = \Sigma_{i=1}^{I} w_i log[\dfrac{1}{1+\exp(-\phi^T x_i)}] + \Sigma_{i=1}^{I}(1-w_i)log[\dfrac{\exp(-\phi^T x_i)}{1+\exp(-\phi^T x_i)}]$
- Gradient : $\frac{\partial L}{\partial \phi} = -\Sigma_{i=1}^{I}(sig[a_i] - w_i)x_i$
- Hessian : $\frac{\partial^2 L}{\partial \phi^2} = -\Sigma_{i=1}^{I} sig[a_i](1 - sig[a_i])x_i x_i^T$

3.3 Results

The optimal values obtained on one trial of the experiment are :(Note : Since the points are sampled randomly, each time we get different optimal points as data changes)

- Optimal ϕ : $\phi = \begin{bmatrix} 0.31 \\ 2.46 \end{bmatrix}$
- Gradient at Optimal ϕ : $\begin{bmatrix} -8.63e-05 \\ -0.00026 \end{bmatrix}$
- Hessian at Optimal ϕ : $\begin{bmatrix} 0.0898 & 0.0898 \\ 0.0898 & 0.0898 \end{bmatrix}$

4 Issue with Linear Separability of Data

When the data is linearly separable, there is an incentive for ϕ to get bigger and bigger, to *emphasize* more and more the difference between the two classes. For example, if you double the size of ϕ, then elements in class 1 get bigger log odds and elements in class 0 get smaller log odds. So from the prespective of minimizing the loss function, it is always better to make ϕ bigger and bigger. In this way, ϕ won't converge. One way to deal with the issue is to add regularization to the cost function which penalizes the ϕ vector for becoming too large.

5 Experiment Results

5.1 Confusion Matrix

The confusion matrix obtained is as follows :

	Class 1	Class 0
Class 1	19 (tp)	1 (fn)
Class 0	0 (fp)	20 (tn)

5.2 Required Plots