2413, Machine Learning, Homework 1
Due Date: 16/10/2013
Universität Bern

Question 1 (3 points) Complete the face classification problem attached to this document, providing a functions that performs the training of a logistic regression classifier.

You are required to:

- Implement the Newton-Raphson method for solving the logistic regression training. You are required to implement a function:

```
function [ theta ] = logisticRegressionTrain_YourName
( DataTrain, LabelsTrain, maxIterations )
% Your code here
```

  Where you have to substitute "YourName" with your name and surname in capital letters. Add comments in your code to explain the most important parts of your algorithm.

- Explain the classification results obtained by the classifier.

Hints:

- The training data has labels in the set $\{-1, 1\}$. Remember to convert them to the range $\{0, 1\}$.
- The function we want to maximise is :

$$\log L(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] \tag{1}$$

  where $h_\theta(x^{(i)}) = g(\theta^T x^{(i)})$ and $g(x) = \frac{1}{1+e^{-x}}$. Remember that $x^{(i)} \in \mathbf{R}^{n+1}$ and $\theta \in \mathbf{R}^{n+1}$, while $h_\theta(x^{(i)})$ and $y^{(i)}$ are scalar. m is the number of training samples and $x^{(i)}$ and $y^{(i)}$ are the samples i in the training set.

- the gradient of $\log L(\theta)$ is:

$$\nabla \log L(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - h_\theta(x^{(i)}))x^{(i)} \tag{2}$$

- the Hessian of $\log L(\theta)$ is:

$$\nabla^2 \log L(\theta) = -\frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1 - h_\theta(x^{(i)}))x^{(i)}(x^{(i)})^T \tag{3}$$

Question 2 (2 points) Consider a regression problem, where the target variables and the inputs are related via the equation

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)} \tag{4}$$

where $\epsilon^{(i)}$ is a error term that captures unmodeled effects. Assume that the $\epsilon^{(i)}$ are distributed IID (Independently and identically distributed) according to a **Laplace distribution** with $\mu = 0$ (Do a research on internet to learn about the Laplace distribution, its PDF and CDF, and its properties).

Derive the expression of the **log likelihood** $l(\theta) = logL(\theta) = logp(y|X;\theta))$.