Homework Assignment 4

In this assignment you will implement logistic regression algorithm over given base code.

Requirements:

1. Apply given logistic regression base code to given Social_Network_Ads.csv file from Kaggle.
2. Plot cost function's result for each iteration of gradient descent.
3. What happens if you change test and training data portions of data to different rates? Explain what you think is causing those changes.
4. Compare results with and without using feature scaling. Explain why one is performing better than another. Clearly explain your reasoning.
5. Do you think it is wise to use USER ID as one of the features for training and predictions? Explain your reasoning!
6. Make learning rate very large (>1, <10). What do you see? Explain clearly.

BONUS:

1. Can we replace gradient descent with Normal Equation for logistic regression? Explain your reasoning!

Note: You can use following examples to replace gender strings (categorical values) to numeric values.

```
data['Gender'].replace(0, 'Female',inplace=True)
data['Gender'].replace(1, 'Male',inplace=True)
```

or same thing in a single line:

```
data['Gender'].replace([0,1],['Female','Male'],inplace=True)
```

Deliverables:

1. Your source code including code for all plots (e.g. **YourFullNAme_HW4_CPSC4370.py**)
2. Word or pdf document of your answers to questions including all visualization plots. Name file: **YourFullNAme_HW4_CPSC4370.docx** or **YourFullNAme_HW4_CPSC4370.pdf**

ANS:

Well not likely, only one discriminative method in classification theory, linear regression... (linear discriminant analysis/fischer discriminant are generative, and even they have a closed form solution due to the extreme simplicity of the distributions fitted).
So, what made Normal Equation so successful in linear regression? Because once you've computed your derivatives, you'll find that the outcome is a set of linear equations, m equations with m variables, which we know can be solved directly using matrix inversions (and other techniques). When logistic regression costs are differentiated, the resultant issue is no longer linear... it is convex (thus global optimum), but not linear, and as a result, present mathematics does not offer us with tools powerful enough to identify the optimum in closed form solution