Assignment 2

Due Date: 10/10/2017

Total Points: 100

In this exercise, you will implement logistic regression and get to see how it works. You will implement the Gradient Descent Algorithm that we have discussed in class to find out the parameters for $\ominus$. A good way to verify that gradient descent is working correctly is to look at the value of J($\ominus$) and check that it is decreasing with each iteration. For implementing some of the principles of programming, try to modularize the code as much as possible and consider **testing** your algorithm on a smaller known dataset before starting the assignment. Please note, that you can also use datasets such as here: http://data.princeton.edu/wws509/datasets to test your algorithm before testing it on the noisy flu dataset. Assignment 2 contains three sections. Please address the subparts in **each** section to receive full credit. Also, analysis is a crucial aspect of the assignment, so for each subpart try to answer the question in more detail.

1. Logistic Regression with One Variable
    a. Can you map the Risk (Risk) of contracting influenza (x) to Flu (y) i.e. whether the student reported flu-like symptoms in the past year?
    b. **Evaluate performance using a metric discussed in class (such as confusion matrix). You may also use graphs for explaining your observations.**

2. Logistic Regression with Multiple Variables
    a. Can you map the Risk and HndWshQUal to Flu (y)?
    b. Does adding the KnowlTrans parameter as another input improve the performance of the model (3 input variables)?
    c. Does adding the Gender improve the performance of the model (so now 4 input variables)? **Evaluate performance for each case using a metric discussed in class (such as confusion matrix).**

3. Regularization and Feature Scaling:
    a. For the best performing model in Q 2, does regularization improve the performance?
    b. Does Feature Scaling improve the performance for the model in Q 3a?
    c. **Evaluate performance for each case using a metric discussed in class (such as confusion matrix).**

Please make sure to submit a zipped file in Dropbox on Pilot titled Assignment 2 with the report in pdf format.

Academic Integrity

Discussion of course contents with other students is an important part of the academic process and is encouraged. However, it is expected that course programming assignments, homework assignments, and other course assignments will be completed on an **individual** basis (unless specified otherwise). Students may discuss general concepts with one another, but may not, under any circumstances, work together on the actual implementation of any course assignment. If you work with other students on "general concepts" be certain to acknowledge the collaboration and its extent in the assignment. Unacknowledged collaboration will be considered dishonest. "Code sharing" (including code from previous quarters) is strictly disallowed. "Copying" or significant collaboration on any graded assignments will be considered a violation of the university guidelines for academic honesty.

If the same work is turned in by two or more students, all parties involved will be held equally accountable for violation of academic integrity. You are responsible for ensuring that other students do not have access to your work: do not give another student access to your account, do not leave printouts in the recycling bin, pick up your printouts promptly, do not leave your workstation unattended, etc. If you suspect that your work has been compromised notify me immediately. If you have any questions about collaboration or any other issues related to academic integrity, please see me immediately for clarification. In addition to the policy stated in this syllabus, students are expected to comply with the Wright State University Code of Student Conduct (http://www.wright.edu/students/judicial/conduct.html) and in particular the portions pertaining to Academic Integrity (http://www.wright.edu/students/judicial/integrity.html) at all times.