**Machine Learning Programming Assignment 6**

**Comp540 Spring 2015 due 17 April 2015 at 8 pm**

**Net ID:MW56&Hz32**

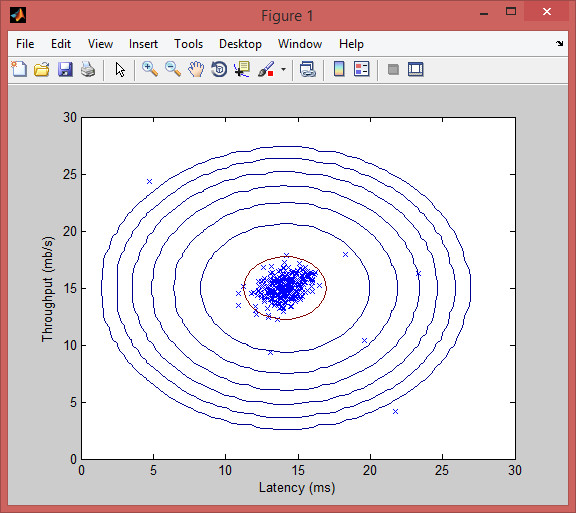
**Miao Wang & HaoYue ZHANG**

**Problem 1: Anomaly detection (30 points)**

**Gaussian distribution**

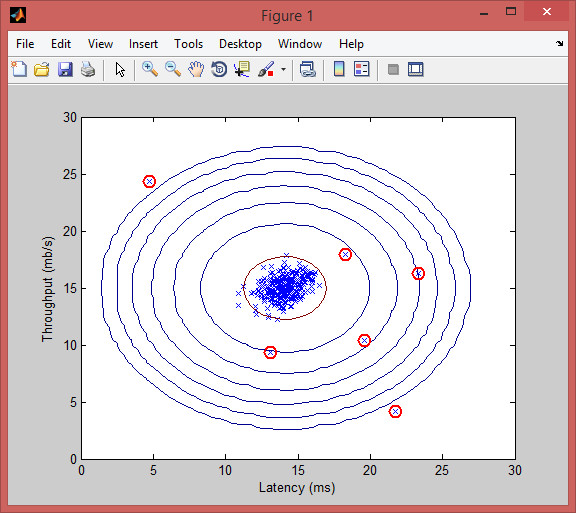
**Estimating parameters of a Gaussian distribution (15 points)**

You should get a plot similar to Figure 2. From your plot, you can see that most of the examples are in the region with the highest probability, while the anomalous examples are in the regions with lower probabilities.

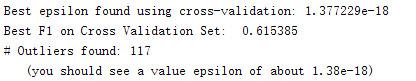


**Selecting the threshold** *E* **(15 points)**

Once you have completed the code in selectThreshold.m, the next step in ex6.m will run your anomaly detection code and circle the anomalies in the plot as in Figure 3.



**High dimensional dataset**



**Problem 2: Recommender systems (70 points)**

Movie ratings dataset

Collaborative filtering learning algorithm

Collaborative filtering cost function

Collaborative filtering gradient

Regularized cost function

Regularized gradient

Loading movie ratings dataset.

Average rating for movie 1 (Toy Story): 3.878319 / 5

Cost at loaded parameters: 22.224604

(this value should be about 22.22)

Checking Gradients (without regularization) ...

5.5335 5.5335

3.6186 3.6186

5.4422 5.4422

-1.7312 -1.7312

4.1196 4.1196

-1.4833 -1.4833

-6.0734 -6.0734

2.3490 2.3490

7.6341 7.6341

1.8651 1.8651

4.1192 4.1192

-1.5834 -1.5834

1.2828 0

-6.1573 0

1.6628 0

1.1686 0

5.5630 0

0.3050 0

4.6442 0

-1.6691 0

-2.1505 0

-3.6832 0

3.4067 0

-4.0743 0

0.5567 0

-2.1056 0

0.9168 0

The above two columns you get should be very similar.

(Left-Your Numerical Gradient, Right-Analytical Gradient)

If your backpropagation implementation is correct, then

the relative difference will be small (less than 1e-9).

Relative Difference: 0.383412

Cost at loaded parameters (lambda = 1.5): 31.344056

(this value should be about 31.34)

**Recommendations**

New user ratings:

Rated 4 for Toy Story (1995)

Rated 3 for Twelve Monkeys (1995)

Rated 5 for Usual Suspects, The (1995)

Rated 4 for Outbreak (1995)

Rated 5 for Shawshank Redemption, The (1994)

Rated 3 for While You Were Sleeping (1995)

Rated 5 for Forrest Gump (1994)

Rated 2 for Silence of the Lambs, The (1991)

Rated 4 for Alien (1979)

Rated 5 for Die Hard 2 (1990)

Rated 5 for Sphere (1998)

Top recommendations for you:

Predicting rating 5.7 for movie One Night Stand (1997)

Predicting rating 5.5 for movie Santa with Muscles (1996)

Predicting rating 5.4 for movie Entertaining Angels: The Dorothy Day Story (1996)

Predicting rating 5.4 for movie Pather Panchali (1955)

Predicting rating 5.2 for movie They Made Me a Criminal (1939)

Predicting rating 5.2 for movie Wild Bunch, The (1969)

Predicting rating 5.2 for movie Prefontaine (1997)

Predicting rating 5.1 for movie Microcosmos: Le peuple de l'herbe (1996)

Predicting rating 5.1 for movie Oscar & Lucinda (1997)

Predicting rating 5.1 for movie Saint of Fort Washington, The (1993)

Original ratings provided:

Rated 4 for Toy Story (1995)

Rated 3 for Twelve Monkeys (1995)

Rated 5 for Usual Suspects, The (1995)

Rated 4 for Outbreak (1995)

Rated 5 for Shawshank Redemption, The (1994)

Rated 3 for While You Were Sleeping (1995)

Rated 5 for Forrest Gump (1994)

Rated 2 for Silence of the Lambs, The (1991)

Rated 4 for Alien (1979)

Rated 5 for Die Hard 2 (1990)

Rated 5 for Sphere (1998)

**Recommendations**

After the additional ratings have been added to the dataset, the script will proceed to train the collaborative filtering model. This will learn the parameters X and Theta. To predict the rating of movie *i* for user *j*, you need to compute (*θ*(*j*))*T x*(*i*). The next part of the script computes the ratings

for all the movies and users and displays the movies that it recommends (Figure 4), according to ratings that were entered earlier in the script. Note that you might obtain a different set of the predictions due to different random initializations.

**What to turn in**

Please zip up all the files in the archive (including files that you did not modify) and submit it as pa6 netid.zip on Owlspace before the deadline. Include a PDF file in the archive that presents your plots and your discussion of results from the problems above.

**Acknowledgment**

These problems are from Andrew Ng’s exercise on anomaly detection and collaborative filtering.