COGS 118A Natural Computation I - Assignment 4

Instructor: Prof. Zhuowen Tu
TA: Chen-Yu Lee
Due: 2/20/2014 11:59 PM

Note: Please send a **PDF** version to tbrsun@gmail.com with the subject [COGS 118A Assignment 4] by due date. No hard copy this time.

1. Normal distribution experiment. Please do the following procedures and show the results in your report:

1. Use "randn" function in MATLAB to generate 1000 1-D samples from normal distribution and name it as variable X.
2. Use "hist" function to compute the histogram of the samples and paste it to the report.
3. Generate another 1000 1-D samples again and name it as variable Y.
4. Plot the scattered points (X,Y) in a 2-D figure using "scatter" function and paste the figure in the report.
5. Now compute two new variables $X' = 3X + Y$ and $Y' = X - 2Y$ and plot the scattered points (X',Y') in a 2-D figure and paste it in the report.
6. Compute the histogram again for X' and Y' and paste the figures in the report.
7. Explain what you see for this experiment from histograms and scattered points.

2. Please download the file "data.mat" from the course website, and load it into your MATLAB environment. Now do the following procedures and show the result in your report:

1. Plot the data using plot function.
```
>> plot(x,y); grid;
```
2. Create the matrix
```
>> A = [x.^0 x.^1];
```
3. Compute the least squares line over the given data by:
```
>> sol = inv(A'*A)*A'*y;
```
4. Overlay the computed least square line over the given data:
```
>> hold on; plot(x, sol(1)+sol(2)*x, '--');
```
5. Assign a title to this figure:
```
>> title('Least squares line fitting'); xlabel('x'); ylabel('y');
```
6. Copy and paste the figure in your report.

3. Now use the same data as in Q2 but compute the least square parabola (i.e. second order polynomial such as $y = ax^2 + bx + c$) to fit the data using similar technique as in Q2.

1. Derive the formulation for the second order least square problem.
2. Write the code to compute the derived parabola using given data, and overlay it on the data point in the same figure as in Q2. Please paste your code and the figure in the report for this question.
3. Explain which formulation (line and parabola) is more suitable for this dataset and why.

4. **(Bonus)** Similar to Q3, but now we are given different weights for each data point when computing the least square error:

$$\text{error} = \sum_{i=1}^{N} t_i(y_i - f(x_i))^2$$

where t_i are weights for each data point (x_i, y_i), and f is the function by design. Please repeat each step in Q3 but now include weights into the error function. For t_i you need to download pre-genereated weights file "weights.mat" on the course website. Show your derived formulation, computed curve figure, and your code in your report.

5. **(Bonus)** Similar to Q3, but now we use L1 norm as your loss function:

$$\text{error} = \sum_{i=1}^{N} |y_i - f(x_i)|$$

where (x_i, y_i) are the given data points, and f is the function by design. Show your approach, computed curve, and your code in your report.