

# CogSci 109 - Homework 7/takehome final section

C. Alex Simpkins

December 8, 2007

## 1 Description

200 pts total

(+15pts extra credit possible)

This homework consists of the final homework and short answer portion of the final exam wrapped into one.

Please read this entire document before proceeding.

### 1.1 Formatting (10 points)

- Please include all matlab code when you turn in your responses. The code may be in an appendix at the end of the assignment, but is preferably included with the problem to facilitate grading (2 pts.)
- Create a cover page which includes your name, the date, the course, instructor, and the course number/title of the assignment. (2 pts.)
- Printouts may be on either new paper or recycled paper, but the recycled paper must have material printed or written on the back which does not distract from the present material (i.e. no bleed-through of ink, no tears, marks, etc). Lined paper may not be used (1 pt)
- Pages must be numbered (1pt.)
- All figures must have axis labels and a figure title. If there is more than one data line or there is a scatterplot of \*s and lines, you must use a legend to clarify which line corresponds to what information source. (2 pts.)

- If you have plots with more than one line, and since you may be printing in black and white (you may also print in color), you must use more than one linestyle to set the lines apart, even if they have different colors. (2 pts.)
- Plots DO NOT have to be in color, but CAN be in color.
- Printing or other problems should be reported immediately so they can be resolved.

## 2 Problems

### 2.1 Problem - Basic data manipulation and loading (30 pts total)

Your friend is working in the new CogSci Virtual Reality Laboratory (CSVRL) and has approached you with some data to analyze. The file is named **CSVRLData.mat**. He or She asks you to help access and visualize the data.

#### 2.1.1 Question (6 points - brief answer)

What kind of data file is it (binary, ASCII)? What is the difference between these formats?

#### 2.1.2 Question (4 points - brief answer)

Can the data in this format (.mat) be easily accessed by Microsoft Word or another text editor without modification? Why?

#### 2.1.3 Question (10 points - brief answer)

How can you access (look at and manipulate) the data in the file (i.e. what are the specific commands and program you might use which we learned in this course)?

#### 2.1.4 Question (10 points - brief answer)

Assuming you have loaded the data into Matlab, you have a matrix of values in memory which are of size 1000x2. One column of data represents x position of a human hand, while

the other column represents y position. The hand moves in a plane parallel to the ground. You want to somehow 'look' at the data. What is one way you can do this (there is more than one possible answer here)?

## 2.2 Problem - Basic data manipulation, visualization and statistics (40 points total)

### 2.2.1 Load the data - (10 points)

Using matlab, download the data file linked on the web site named **finaldata.zip**. Expand the file and you should have a folder called finaldata. Inside that folder there is a file called:

**problem2.mat**

Load this file into matlab using a command or technique you learned in this class, just like all the homeworks where you loaded data into matlab from a data file.

List the matlab commands you use to do this (10 points).

### 2.2.2 Plot the data - (10 points)

Plot this data so that each variable in memory is plotted in a separate figure window. Each figure should have a title, xlabel and ylabel (worth 3 points towards plot points, one point each component - ie title is worth one point, xlabel worth one point, etc). There should be two plots (5 total points per plot).

### 2.2.3 Compute the basic statistics - (10 points)

Write down all values from this section in a table (which can be hand-written, created in word, excel, etc) and include them in your writeup. Include the code to perform the computations in matlab.

- Write down the equation for the sample mean (2 points), and compute the mean of each data variable (there should be two means, one for each variable in the **problem2.mat** file). (3 points).

- Write down the equation for the sample standard deviation (2 points), and compute the standard deviation of each data variable (there should be two standard deviation's, one for each variable in the **problem2.mat** file) - (3 points).

#### 2.2.4 General interpretation - brief answer (10 points)

In general, what does the concept of the mean tell us about our data (4 points)? What does the concept of the standard deviation tell us (4 points)? In this specific instance, what is the numerical relationship between the value of the mean of subject1 vs. subject2, and what is the numerical relationship between the value of the standard deviation of subject1 vs. subject2 (2 points)? (Hint - calculate and compare)

### 2.3 Problem - Choosing a fit (40 points)

You are given the following data (you can copy and paste this into matlab):

$$x = [0 \ 2 \ 4 \ 6 \ 8 \ 10]$$

$$y = [-1.4056 \ -0.3855 \ 1.9451 \ 20.5890 \ 40.0299 \ 70.4374]$$

Using a method of your choice, create a fit for this data.

#### 2.3.1 Question - (10 points)

Choose a fit method which was discussed in this course (for example, linear least squares, nonlinear least squares, lagrange, splines, LERP, artificial neural network). In one or two sentences explain the motivation for your choice. Note: There are multiple acceptable answers.

#### 2.3.2 Question - (10 points)

Perform the fit. Include your matlab code.

#### 2.3.3 Question - (10 points)

Plot your curve on the same plot as the raw data. Include a title, legend and axis labels. The data and fit curves should be clearly distinguishable.

### 2.3.4 Question - (10 points)

Perform a simple error analysis by computing the sum of the squared error, norm-based-error, or plotting the difference error ( $e=y_{data}-y_{fit}$ ). Comment on your fit's quality/good or bad sections of the fit/etc (briefly).

## 2.4 Problem - data modeling with Artificial Neural Networks (40 points)

Consider the following (imaginary) case. You have recorded data from two variables in language comprehension. You want to fit a curve to that data, then use the fit to predict future outputs given particular inputs (i.e. you want to derive a relationship between the variables).

- If you have not yet done so, download the file **finaldata.zip**
- Load the data file **problem3.mat**

You should have these two variables from this file - P and T. P represents the inputs, T will be the training set to relate the inputs to the outputs

### 2.4.1 Question - (10 points)

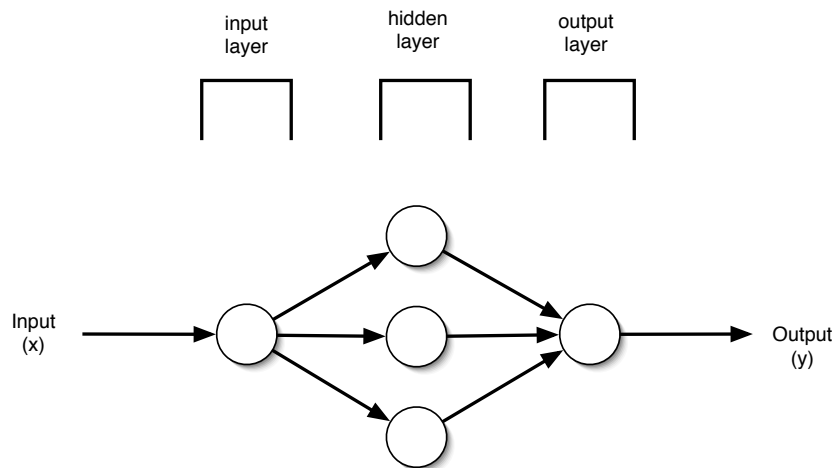
Draw a basic feedforward system block diagram as shown in class (3 points), label the key components and signals

Draw a basic feedback system block diagram as shown in class, label the key components and signals (3 points). What is the fundamental difference between the two systems (4 points)?

### 2.4.2 Question - (10 points)

Let's model the given data with a multilayer feedforward artificial neural network (a feed-forward backpropagation network).

We have one input, let us select 1 hidden layer with 3 nodes, and one final output (3 nodes should be sufficient to fit this function, found through trial and error). The network can be represented visually in the following way:



If you have access to the matlab neural network toolbox, download and run the code **problem3.m** from the **finaldata.zip** folder. This generates data and results for problem 3. If you have any trouble getting the network to converge, be sure to perform a clear all at the command line before running the code. If you don't have access to the neural network toolbox in matlab, the results of the code are provided in **problem3.mat**

Plot the results (actual data, P vs. T as small circles, stars, or another shape, the fit, P vs. Y as a solid line) with a legend, title (associated with the problem number), xlabel and ylabel (**10 points**).

### 2.4.3 Questions - (20 points)

Given the matlab command line output as follows (these are result listings from running problem3.m):

```
TRAINLM-calcjx, Epoch 7/2000, MSE 0.000998381/0.001, Gradient 0.0266953/1e-10
```

```
TRAINLM, Performance goal met.
```

Answer the following:

- (5 points) Write down the equation for mean-squared-error
- (5 points) Explain the parts of the mean-squared-error equation
- (10 points) What is the purpose of having a criterion such as mean-squared-error (i.e. what does it tell us or the algorithm we're using)?

## **2.5 Problem - how to approach problems (40 points)**

You have run an experiment on adaptation in human sensorimotor control. Your goal is to write a paper discussing a robotic simulation of the human subjects' movement, with a demonstration that certain consistent perturbations (force fields) applied to the subject will result in an adaptation. You have by this point massive quantities of data, and are attempting to model the data with no prior knowledge of the physical relationships between variables (the relationships are obviously complex and subtle). Given what you know about neural network models, you decide to try to model the data with an artificial neural network. Please answer the following:

### **2.5.1 Question - (20 points)**

What are the four steps of modeling and data analysis as presented in this course (write them down, and give one or two sentence description)?

### **2.5.2 Question - (20 points)**

Given those four steps, how could you use them to address this problem? Take each step, and discuss in roughly one sentence per topic a breakdown of this modeling problem from beginning to end.

## **3 Bonus - (9 points)**

Considering the entire course, all that you have performed in assignments, prior experience, readings, and any additional information you have acquired, it is important to try to pull it together. Therefore, in your own words, please respond to the following (very briefly):

(+5 points) What does data analysis mean?

(+5 points) What does modeling mean?

(+5 points) How do the above two topics relate to Cognitive Science (or your related field)?  
Stating that there is no relationship does not constitute an acceptable answer.

**[END OF HOMEWORK/TAKEHOME SECTION]**