Statistics and Data Analysis in MATLAB

Kendrick Kay, kendrick.kay@wustl.edu

Homework 3 (covering Statistics Lectures 3 and 4)

To complete this assignment, prepare a MATLAB script called `homework3.m` along with any necessary accompanying function .m files. Then, run the MATLAB `publish` command (e.g. `publish('homework3.m')`) to run the script and generate HTML output showing the results. Turn in a print-out of the HTML output (e.g. from your web browser) and also a print-out of any function .m files that you write.

Hint: In your script file, place `%%` on a line by itself at each point where you want the HTML output to show figures and command-window text. Please note that your code should be commented (where necessary), including documentation of any functions that you write.

Problem 0. Download the .mat file at http://artsci.wustl.edu/~kkay/psych5007/Homework3.mat (you will need this file to complete the problems below).

Problem 1. The `score1` and `score2` variables in Homework3.mat contain scores for 100 subjects on two different tests. Fit a linear model ($y = ax + b$) to these data, attempting to predict `score2` from `score1`. Visualize the data using a scatter plot and visualize the fitted model using a line. Report the estimated parameters in the figure title.

Problem 2. Redo problem 1, this time using a quadratic model ($y = ax^2 + bx + c$).

Problem 3. Redo problem 1, this time using the model $y = ax^n + b$. Start from two different initial seeds: $a = 1$, $n = 1$, $b = 0$ and $a = 10$, $n = 7$, $b = 100$. Plot the model that achieves a lower squared error using a regular line and plot the other model using a dashed line.

Problem 4. Perform gradient descent to fit the model described in Problem 1 (see the lecture notes for Stats Lecture 4). (Hint: The amount by which to update the weights is a tricky issue. After computing the gradient vector, normalize it to have a length of one. Then update the weights by a small fraction of this vector.) Report to the command window the parameter estimates obtained by your gradient descent routine. Confirm that they are more or less identical to the parameter estimates you obtained in Problem 1.