HW 6: Linear Regression Questions 29-36

Important rules:

- You are not allowed to use statistical packages! You need to implement linear regression yourself, because it helps you understand how the different quantities involved are related.
- As usual the collaboration policy will be strictly enforced. You are not allowed to solve this assignment together or share code or take code from the internet!
- Do not use blackboard as a debugging platform! In fact, you won't get any feedback from Blackboard. We provide you with example data to test your implementation.

Important formulas for linear regression:

- Simple linear regression: $a = \frac{Cov(X,Y)}{Var(X)}$ and $b = E[y] - E[ax]$.

  These equations yield the parameters for the regression ax+b=y. They are valid for one covariate (only one x) and correspond to a minimization of mean squared error.
- Correlation coefficient: $r = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$
- (For people with statistics background: We are treating our data as a population not a sample.)

Debugging:

- The expected outputs for your implementation are:
  - debug1.txt: a = 2.1885 (2.2) b = 4.7503 (5) r = 0.9411 (0.9)
  - debug2.txt: a = 0.2109 (0.2) b = 6.4350 (6) r = 0.8679 (0.9)
- All values are rounded not truncated. (values in parenthesis indicate rounding for blackboard)

Plotting:

- You are allowed to plot with whatever you want (e.g. in Excel you can create a scatter plot)
- If you don't know what to use you can use gnuplot http://sourceforge.net/projects/gnuplot/files/gnuplot/4.6.6/
- The basic plotting command in gnuplot is "plot <filename> (with <plotstyle>)", it expects the data in the exact same format as provided. You can also plot functions by defining a function, e.g. "f(x) = 10*x+4" and then "replot f(x)" (replot adds it to the already displayed plot)

Administrative Issues:

- Compress all of your code in a single tgz file. Do not compress the directory that contains the files but the files themselves. You can use e.g. tar –czvf hw6.tgz *.java
- Don't forget to fill out the collaboration form on blackboard.
- Don't forget to **submit** your homework before the deadline! (not just save)
- The deadline is Tuesday, 02/24/2015 23.59pm EST, you get exactly 3 submissions and you will not get any feedback until after the deadline.