CIS/EE 794R/694R **Spring 2004**
Prof. Paul Sivilotti

Assignment #4: A JSP Datamining Application

DUE: 10:30 am, Friday, May 14th.

My daughter is preparing for her 4th grade science fair project. For this project, she is investigating the factors that affect how bouncy a ball is. That is, when a ball is dropped from a height H, what is the height (as a percentage of H) to which it bounces on the first hop?

For this experiment, she has fixed: (i) the surface on which the balls will be bouncing (*i*.e., the kitchen floor), and (ii) the material of construction of the balls. Measurements are taken of different balls (masses and colors) falling from different heights, at different temperatures.

The gathered data is contained in a text file (see newsgroup posting for location).

Part I: JSP Linear Regression Engine

Multivariable linear regression calculates the best-fit hyperplane through a multidimensional space. That is, for a collection of data with one dependent variable, y, and a set of independent variables, x_1, x_2,..., x_n, it calculates the coefficients of the equation:

$$y = c_1 * x_1 + c_2 * x_2 + ... + c_n * x_n$$

This hyperplane is "best fit" in the sense that the coefficients result in the smallest sum-of-squares difference between the equation and the actual data.

Write a JSP multivariable linear regression engine for the bounciness dataset. The engine should support analysis of a subset of the experiments, based on a variety of criteria. For example, the user might wish to:

- run the regression only those data points within a certain range of temperatures,
- compare the regression results if only blue balls are considered versus only green balls,
- list the 10 trials which resulted in the highest bounciness

Note that these are only examples of the kinds of querries the user might wish to express.

You will develop your JSP pages in a subdirectory of

```
/usr/class/cis894/jakarta-tomcat-5.0.19/webapps
```

You will receive email from the grader indicating the name of the directory you should use. Please note the following important points:

- *Do not share* the name of this directory with anyone else in the class. Your JSP files will be world readable so you should protect your solutions by keeping the directory name confidential.
- *Do not modify* the contents of this directory after the lab due date. We will consider the contents of this directory to be your "submission", so the timestamp of the files should predate the lab due time.
- Include a index.html file in this directory so that we can test your system without knowing the name of a particular page in your directory.
- Please confine your solutions to JSP. Reading in the text file and storing the data in memory, for example, is preferred (for the pursposes of this assignment!) to creating a database.

Part II: Front-end for Analysis

Using Javascript, write a front-end for your linear regression engine. Your goal is to deploy an interface that allows us to determine the relationships among these various factors. Thus, your interface must be usable and flexible, allowing us to issue a variety of "useful" querries for understanding these relationships. Of course your site should contain information that helps the user understand your interface.

In addition to the code, for this lab you should also submit a written analysis of the data (carried out via your user interface!) and conclusions about the co-efficients for the various terms.