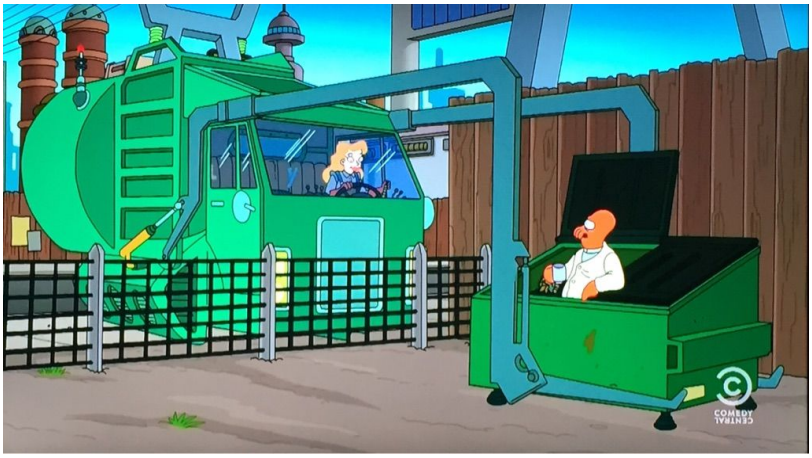
HOMEWORK #1:

Linear Regression and Trashcans

Due Date: Thursday, October the 12th, 11:59:59pm

Problem:

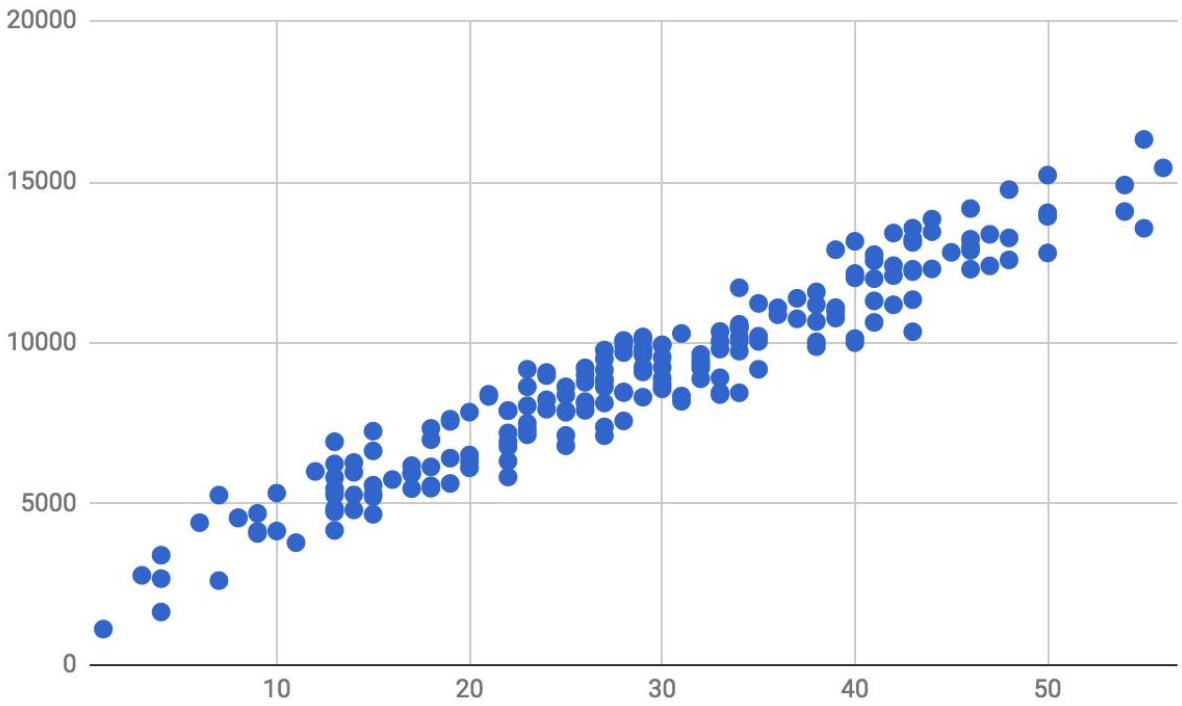
Dr. Zoidberg is always in the lookout for a good meal, especially if it is free! And the best meals can be found in the trash cans behind restaurants. Dr. Zoidberg has surveyed the city's restaurants and over many months has collected extensive data on the number of customers a restaurant serves during the day together with the quantity of tasty treats that can be found in the restaurant's trash can at the end of the day. Help Dr. Zoidberg make predictions about how much he can expect to eat given the number of customers a restaurant serves. You will train a "single neuron" to perform regression on the data collected by Dr. Zoidberg.



Getting food this way also has risks

Input:

The given file `zdata.txt` contains the data for this regression problem. The first column is the number of customers a restaurant serves in a particular day. the second column is the number of calories that Dr. Zoidberg found at the end of the day. You can visualize the data in the following scatter plot:



Gradient Descent:

You will implement gradient descent, and will use Least Squares as the error measurement for your neuron. Your objective is then to minimize the function:

$$LLE(E) = \Sigma_{e \text{ in } E} (Y(e) - \hat{Y}(e))^2$$

where E is the set of examples, $Y(e)$ is the data value for an example e, and $\hat{Y}(e)$ is the output of the neuron given by

$$\hat{Y}(e) = \Sigma_{i=0}^{n} (W_i * X_i(e))$$

Notice that we are not applying a squash function to this neuron. This is because we are doing *regression*, and not *classification*.
You shall implement **batch** gradient descent. This means that you will be adding the errors for all the examples and then updating the neuron's weights. Your program should perform **1500** iterations of the gradient descent. Initialize the weights randomly. The learning rate will be fixed at η **= 0.01**.

Submission Guidelines:

You will submit through the department's Unix machines using the command:

```
cssubmit 5001 a 1
```

Your submission will consists of the following components:

1. Your program files.- Submit all necessary files. Your program should compile and run in the Unix systems. Your program should read the input data from a file called '`zdata.txt`', (in the same location as your program), and produce output formatted like in the sample shown below.
2. A `make` file that should build and run your program.
3. A report, in PDF format, in which you state the final weights of your neuron, and plot your regression formula against the data.

Sample Output:

Output

```
CS-5001 : HW#1 : Regression with one variable.
Programmer: Dr. Hubert J. Farnsworth

Learning rate eta = 0.01
After 1500 iterations:
Sum of Errors = 0.1234
Weights:
w0 = 3.1415
w1 = 42.0
```

Pseudocode:

```
PROCEDURE RegressionOneInput
        read data into examples array of <x,y> pairs
        randomize w0, w1
        Local: array or errors yErr
        Repeat 1500 times
                For every <x,y> in examples  //Computing errors//Haha! need indexes.Python fail
                        Compute yCap ( from w0, w1 and x )
                        yErr[] := y - yCap
                For every <x,y> in examples  //Updating weights
                        update wi using yErr[] and x
        done? did not got lost? output w0, w1, and sum of yErr[]
```

END.