More Gradient Descent Questions

A single feature can be left alone as a predictor, but sometimes it is useful to manipulate the feature to help it be better used in a linear model.

In the following three questions, the left column represents a feature and the right column represents the current known output. What would be the best way to manipulate the feature, if any, before using it to predict the output?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.   |  |  | | --- | --- | | x | y | | 5  33  65  78  139  171  262  263  398  403  407  431  445  461  476 | 1116  787  3376  4054  3975  5134  5825  5482  9271  10336  9225  10891  10681  10191  9859 | | 2.   |  |  | | --- | --- | | x | y | | 5  33  65  78  139  171  262  263  398  403  407  431  445  461  476 | 127  300  415  443  603  662  816  823  1017  1008  1015  1061  1056  1090  1092 | | 3.   |  |  | | --- | --- | | x | y | | 5  33  65  78  139  171  262  263  398  403  407  431  445  461  476 | 66  22  61  66  276  525  1825  1841  6317  6569  6800  8077  8869  9864  10843 | |

No Change Square root x Cube x

4. You run gradient descent for 15 iterations with *α*=0.3 and compute *J*(*θ*) after each iteration. You find that the value of *J*(*θ*) **decreases** quickly then levels off. What conclusion should you make of your *α* value?

That means the alpha aggression value works well

5. Suppose you have *m*=28 training examples with *n*=4 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is:



For the given values of *m* and *n*, what are the dimensions of *θ*, *X*, and *y* in this equation?

*Θ = 1 x 4*

*X = 28 x 4*

*Y = 28 x 1*