Gradient Descent Review

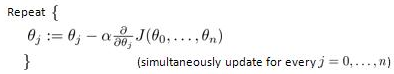
Hypothesis function:



Cost function:



Gradient descent:



Hypothesis function:

1. Describe the purpose of a ‘hypothesis function’.

**A hypothesis function is a prediction that serves to test values.**

1. What do the θ values mean? What is the difference between θ and θ0 / θ1 / θn?

**Thetas are the coefficients of the corresponding x’s. Theta of 0 will always be the bias (y-intercept). Theta of 1 is the first coefficient. Theta of n is the nth coefficient.**

1. What does the T mean in θT?

**The T represents the fact that the matrix with the thetas is “flipped”. (Tranposed)**

1. What is the difference between x and x0 / x1 / xn?

**The x’s are the actual test values. X of 0 is always one. X of 1 is x. x of n is x to the nth power.**

1. What is the value of x0?

**One.**

Cost Function:

1. Describe the purpose of a ‘cost function’.

**It tells you how far off your hypothesis function guesses are.**

1. What does the m mean?

**The amount of inputs.**

1. Describe the values being summed together.

**The y value of the x’s your hypothesis function creates minus the actual y’s, each squared and added to the sum.**

1. If gradient descent is correctly done, what is the ultimate goal of the cost function?

**To get as close to zero as possible.**

1. Gradient Descent:
2. What does α mean?

**Alpha is the aggression or learning rate in machine learning (It’s really a derivative, positive or negative). Generally, between 0-3. Too high, and you might skip past the lowest value. To low, and the calculations take too long.**

1. What should be done if the plot of the cost function values between gradient descent iterations is decreasing quite slowly for a while then decreases even less later?

**Multiply alpha by 3.**

1. What should be done if the plot of the cost function values between gradient descent iterations is decreasing quite quickly for a while then slows down?

**Keep it.**

1. What should be done if the plot of the cost function values between gradient descent iterations is increasing?

**Start in a different spot. (Change the hypothesis function)**

Other questions:

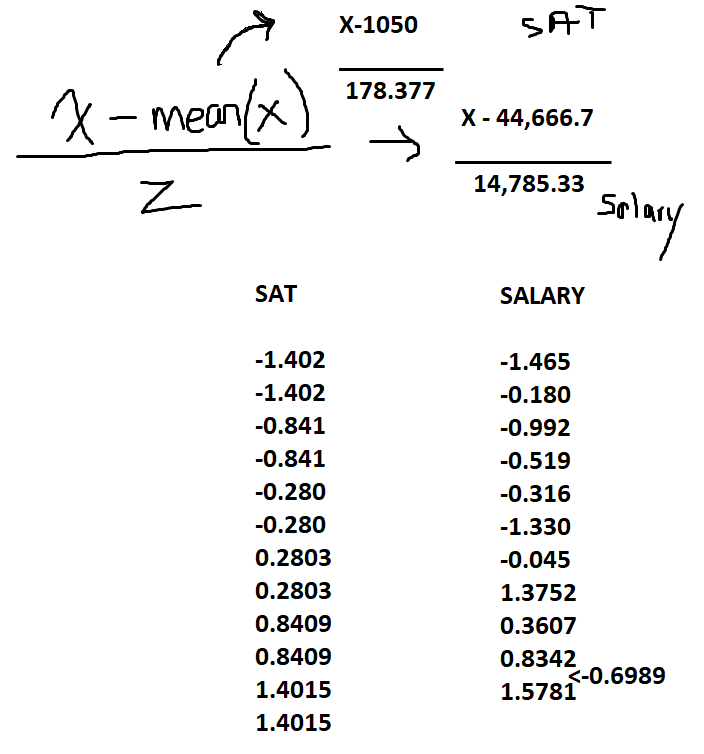
1. What is a ‘feature’?

**I single thing used to calculate the answer. There can be variable amounts of features in data sets.**

1. What should you do to an entire set of features to allow feature to be better balanced against others?

**Standardize the data set.**

A study is being done to determine a model that mimics SAT score results from high school with the median salary a person should expect. The following is some data about SAT scores of individuals (1600 scale) and the salary they made after 10 years of working (everything has been rounded to make it easier).

SAT Salary

800 23000

800 42000

900 30000

900 37000

1000 40000

1000 25000

1100 44000

1100 65000

1200 50000

1200 57000

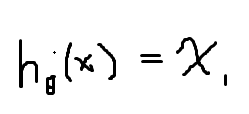
1300 55000

1300 68000

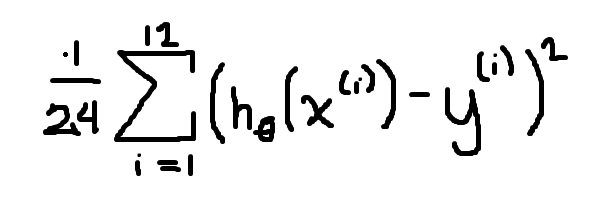
1. How many features do you have?

**1.**

1. Scale the feature(s) using standardization.
2. Write a non-vectorized hypothesis function which begins with θ1 as 1 and θ0 as 0.



1. Calculate the cost function of this hypothesis.



**=1,097,739,397**

1. Using an α value of 0.3, calculate the next hypothesis function (only change θ1 for this problem).
2. Calculate the cost function of this new hypothesis – is it better or worse?
3. Repeatedly use the α value of 0.3 and determine if the θ1 value begins converging. Would it be better to use a smaller or larger value of α?
4. Can you find a way to graph the cost function values using a calculator or other software?