Gradient Descent Review

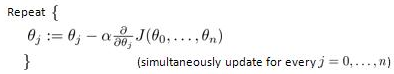
Hypothesis function:



Cost function:



Gradient descent:



Hypothesis function:

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

The purpose of the hypothesis function is to take in multiple variables in a function and simulate data.

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

The theta values represent the “Weight” or coefficients of the different variables. It determines how much each variable is worth in the equation.

1. What does the T mean in θT?

The T in the theta symbol represents a transposed matrix.

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

Each x value represents a different variable factor that influences the function.

1. What is the value of x0?

1

Cost Function:

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

The cost function is used to measure the accuracy of the hypothesis function to the actual verified result.

1. What does the m mean?

The m represents the number of data points in the set.

1. Describe the values being summed together.

Each value being summed together is a value calculated from the cost function given the data point from the index of *i,* minus the actual data point at *i,* with the quantity squared.

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

The ultimate goal of the cost function is to be as close to 0 as possible.

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

The alpha symbol represents the aggression in the learning rate, which is how may units a step is along the graph.

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?

If the values for the cost function are decreasing remarkably slowly, then it would be a good idea to increase the aggression of the learning rate.

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?

If it decreases quickly for a while, and then slows, then that is the ideal behavior for a gradient descent, and data can be collected once the decrease becomes negligible.

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

If the cost function values are increasing, that means that the aggression rate is too high, and it should be lowered.

Other questions:

1. What is a ‘feature’?

A feature is an aspect of a cost function that is deemed important and influential to the calculation. It is a theta value, or coefficient.

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

Normalize the set of features to get them into a similar numerical range.

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?

One feature

1. Scale the feature(s) using standardization

|  |  |
| --- | --- |
| SAT | SAT Standardized |
| 800 | -1.402 |
| 800 | -1.402 |
| 900 | -0.841 |
| 900 | -0.841 |
| 1000 | -0.28 |
| 1000 | -0.28 |
| 1100 | 0.2803 |
| 1100 | 0.2803 |
| 1200 | 0.8409 |
| 1200 | 0.8409 |
| 1300 | 1.4015 |
| 1300 | 1.4015 |

1. Write a non-vectorized hypothesis function which begins with θ1 as 1 and θ0 as 0.

h(x) = 0 + 1x0

1. Calculate the cost function of this hypothesis.

1097705334

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?