

Can someone explain to me the difference between a cost function and the gradient descent equation in logistic regression?

[Ask Question](#)

48



34

I'm going through the ML Class on Coursera on Logistic Regression and also the Manning Book Machine Learning in Action. I'm trying to learn by implementing everything in Python.

I'm not able to understand the difference between the cost function and the gradient. There are examples on the net where people compute the cost function and then there are places where they don't and just go with the gradient descent function $w := w - (\alpha) * (\text{delta})_w * f(w)$.

What is the difference between the two if any?

[machine-learning](#)

edited Dec 15 '16 at 19:50

[alex](#)**350k**

173

784

919

asked Nov 29 '12 at 9:58

[oktapodi](#)**385**

1

6

11

[Home](#)[PUBLIC](#)[Stack Overflow](#)[Tags](#)[Users](#)[Jobs](#)[Teams](#)[Q&A for work](#)

[Learn More](#)

3 Answers



79



A cost function is something you want to minimize. For example, your cost function might be the sum of squared errors over your training set. Gradient descent is a method for finding the minimum of a function of multiple variables. So you can use gradient descent to minimize your cost function. If your cost is a function of K variables, then the gradient is the length- K vector that defines the direction in which the cost is increasing most rapidly. So in gradient descent, you follow the negative of the gradient to the point where the cost is a minimum. If someone is talking about gradient descent in a machine learning context, the cost function is probably implied (it is the function to which you are applying the gradient descent algorithm).

answered Nov 29 '12 at 13:50

**bogatron**

12.1k 2 38 38

- 1 so basically we use the gradient to find out the least cost for the function? – [oktapodi](#) Nov 29 '12 at 16:06
- 2 Correct. To be clear though, you don't use a single gradient value. As you move along the negative direction of the gradient, the gradient will change (unless you are on a hyperplane), so you keep updating the gradient as you move along the direction that most rapidly reduces the cost until you hit a minimum (hopefully, the global minimum). – [bogatron](#) Nov 29 '12 at 17:06

hey, can I have some descent links to learn (like get a basic idea) all these stuff regarding logistic regression, gradient regression and gradient descent? – [Anuroop Pendela](#) Jan 26 '17 at 7:48

- 1 [coursera.org/learn/machine-learning](https://www.coursera.org/learn/machine-learning) – [yegeniy](#) Mar 19 '18 at 4:04



104

Whenever you train a model with your data, you are actually producing some new values (predicted) for a specific feature. However, that specific feature already has some values which are real values in the dataset. We know the closer the predicted values to their corresponding real values, the better the model.

Now, we are using cost function to measure how close the predicted values are to their corresponding real values.

We also should consider that the weights of the trained model are responsible for accurately predicting the new values. Imagine that our model is $y = 0.9 \cdot X + 0.1$, the predicted value is nothing but $(0.9 \cdot X + 0.1)$ for different X s. [0.9 and 0.1 in the equation are just random values to understand.]

So, by considering Y as real value corresponding to this x , the cost formula is coming to measure how close $(0.9 \cdot X + 0.1)$ is to Y .

We are responsible for finding the better weight (0.9 and 0.1) for our model to come up with a lowest cost (or closer predicted values to real ones).

Gradient descent is an optimization algorithm (we have some other optimization algorithms) and its responsibility is to find the minimum cost value in the process of trying the model with different weights or indeed, updating the weights.

We first run our model with some initial weights and gradient descent updates our weights and find the cost of our model with those weights in thousands of iterations to find the minimum cost.

One point is that gradient descent is not minimizing the weights, it is just updating them. This algorithm is looking for minimum cost.

edited Mar 5 '18 at 17:23



[Ravi Kumar Gupta](#)

758 10 30

answered Nov 6 '16 at 1:41



[Reihan_amn](#)

1,787 1 10 15

2 Beautiful answer. :) – [Swaathi Kakarla](#) Jan 12 '17 at 4:55

Reading second time is helping to understand beauty :) – [mert](#)
Feb 16 '17 at 11:53

very well explained . :) – [KMittal](#) Apr 6 '17 at 8:02

Can't get simpler than this ... – [sHiBuKaLiDhAsAn](#) Dec 3 '17 at 12:29

1 To summarize, the gradient descent will find those random 0.9 and 0.1 in order to make our predicted values be as close as possible to the real ones. – [Alisson](#) Sep 21 '18 at 0:27



Cost function is something is like at what cost you are building your model for a good model that cost should be minimum. To find the minimum cost function we use gradient descent method. That give value of coefficients to determine minimum cost function

edited Aug 9 '18 at 11:31

answered Aug 9 '18 at 7:48



[gaurav singh](#)

39 3