derivative of cost function for Logistic Regression

I am going over the lectures on Machine Learning at Coursera.

I am struggling with the following. How can the partial derivative of

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^i \log(h_\theta(x^i)) + (1-y^i)\log(1-h_\theta(x^i))$$

where $h_\theta(x)$ is defined as follows

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}}$$

be

$$\frac{\partial}{\partial\theta_j} J(\theta) = \sum_{i=1}^{m} (h_\theta(x^i) - y^i)x_j^i$$

In other words, how would we go about calculating the partial derivative with respect to θ of the cost function (the logs are natural logarithms):

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^i \log(h_\theta(x^i)) + (1-y^i)\log(1-h_\theta(x^i))$$

(statistics) (regression) (machine-learning) (partial-derivative)

edited Aug 27 '13 at 12:26
Avitus
10.7k 15 37

asked Aug 27 '13 at 10:41
dreamwalker
360 1 4 6

I think to resolve θ by gradient will be hard way (or impossible??). Because it different with linear classfication, it will not has close form. So i suggest you can use other method example Newton's method. BTW, do you find θ using above way? – mjohn1282 Jul 22 '14 at 2:16

1 missing $\frac{1}{m}$ for the derivative of the Cost – bourneli Apr 20 at 5:01

3 Answers

The reason is the following. We use the notation

$$\theta x^i := \theta_0 + \theta_1 x_1^i + \dots + \theta_p x_p^i.$$

Then

$$\log h_\theta(x^i) = \log\frac{1}{1+e^{-\theta x^i}} = -\log(1+e^{-\theta x^i}),$$

$$\log(1-h_\theta(x^i)) = \log(1-\frac{1}{1+e^{-\theta x^i}}) = \log(e^{-\theta x^i}) - \log(1+e^{-\theta x^i}) = -\theta x^i - \log(1+e^{-\theta x^i}),$$

[this used: $1 = \frac{(1+e^{-\theta x^i})}{(1+e^{-\theta x^i})}$, the 1's in numerator cancel, then we used: $\log(x/y) = \log(x) - \log(y)$]

Since our original cost function is the form of:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^i \log(h_\theta(x^i)) + (1-y^i)\log(1-h_\theta(x^i))$$

Plugging in the two simplified expressions above, we obtain

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[-y^i(\log(1+e^{-\theta x^i})) + (1-y^i)(-\theta x^i - \log(1+e^{-\theta x^i}))\right]$$

, which can be simplified to:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \theta x^i - \log(1+e^{-\theta x^i})\right] = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \log(1+e^{\theta x^i})\right], \quad (*)$$

where the second equality follows from

$$-\theta x^i - \log(1+e^{-\theta x^i}) = -\left[\log e^{\theta x^i} + \log(1+e^{-\theta x^i})\right] = -\log(1+e^{\theta x^i}).$$

[we used $\log(x) + \log(y) = log(xy)$]

All you need now is to compute the partial derivatives of $(*)$ w.r.t. θ_j. As

$$\frac{\partial}{\partial\theta_j} y_i\theta x^i = y_i x_j^i,$$

$$\frac{\partial}{\partial\theta_j}\log(1+e^{\theta x^i}) = \frac{x_j^i e^{\theta x^i}}{1+e^{\theta x^i}} = x_j^i h_\theta(x^i),$$

the thesis follows.

edited Sep 17 '15 at 2:32
HRtsFan
3 1

answered Aug 27 '13 at 12:25
Avitus
10.7k 15 37

4 This is terrific, thank you very much! – dreamwalker Aug 27 '13 at 13:31

1 Can't upvote as I don't have 15 reputation just yet! :) Will google the maximum entropy principle as I have no clue what that is! as a side note I am not sure how you made the jump from log(1 - hypothesis(x)) to log(a) - log(b) but will raise another question for this as I don't think I can type latex here, really impressed with your answer! learning all this stuff on my own is proving to be quite a challenge thus the more kudos to you for providing such an elegant answer! :) – dreamwalker Aug 27 '13 at 13:54

1 yes!!! I couldn't see that you were using this property $\log(\frac{a}{b}) = \log a - \log b$ Now everything makes sense :) Thank you so much! :) – dreamwalker Aug 27 '13 at 14:26

3 Awesome explanation, thank you very much! The only thing I am still struggling with is the very last line, how the derivative was made in

$$\frac{\partial}{\partial\theta_j}\log(1+e^{\theta x^i}) = \frac{x_j^i e^{\theta x^i}}{1+e^{\theta x^i}}$$

? Could you provide a hint for it? Thank you very much for the help! – Pedro Lopes Dec 1 '15 at 21:40

2 @codewarrior hope this helps.

$$\frac{\partial}{\partial\theta_j}\log(1+e^{\theta x^i}) = \frac{x_j^i e^{\theta x^i}}{1+e^{\theta x^i}}$$

$$= \frac{x_j^i}{e^{-\theta x^i} * (1+e^{\theta x^i})}$$

$$= \frac{x_j^i}{e^{-\theta x^i} + e^{-\theta x^i + \theta x^i}}$$

$$= \frac{x_j^i}{e^{-\theta x^i} + e^0}$$

$$= \frac{x_j^i}{e^{-\theta x^i} + 1}$$

$$= \frac{x_j^i}{1 + e^{-\theta x^i}}$$

$$= x_j^i * h_\theta(x^i)$$

as

$$h_\theta(x^i) = \frac{1}{1+e^{\theta x^i}}$$

– Rudresha Jan 2 at 13:06

Pedro, => partial fractions

$$\log(1-\frac{a}{b})$$

$$1-\frac{a}{b} = \frac{b}{b} - \frac{a}{b} = \frac{b-a}{b},$$

$$\log(1-\frac{a}{b}) = \log(\frac{b-a}{b}) = \log(b-a) - \log(b)$$

edited Apr 13 '16 at 15:39

answered Apr 13 '16 at 15:23

Richard Wheatley
21 3

@pedro-lopes, it is called as: chain rule.

$$(u(v))' = u(v)' * v'$$

For example:

$$y = \sin(3x-5)$$

$$u(v) = \sin(3x-5)$$

$$v = (3x-5)$$

$$y' = \sin(3x-5)' = \cos(3x-5) * (3-0) = 3\cos(3x-5)$$

Regarding:

$$\frac{\partial}{\partial\theta_j}\log(1+e^{\theta x^i}) = \frac{x_j^i e^{\theta x^i}}{1+e^{\theta x^i}}$$

$$u(v) = \log(1+e^{\theta x^i})$$

$$v = 1+e^{\theta x^i}$$

$$\frac{\partial}{\partial\theta}\log(1+e^{\theta x^i}) = \frac{\partial}{\partial\theta}\log(1+e^{\theta x^i}) * \frac{\partial}{\partial\theta}(1+e^{\theta x^i}) = \frac{1}{1+e^{\theta x^i}} * (0+xe^{\theta x^i}) = \frac{xe^{\theta x^i}}{1+e^{\theta x^i}}$$

Note that

$$\log(x)' = \frac{1}{x}$$

Hope that I answered on your question!

edited Apr 17 at 13:29

The Count
2,221 6 11 29

answered Apr 17 at 13:17

Vadim Stupakov
11 2