Proofs you probably weren't taught

Friday, October 21, 2011

Why is the error function minimized in logistic regression convex?

We want to prove that the error/objective function of logistic regression :

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

$$where \;\; h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

is convex.

Proof:

Before beginning the proof, i would first like to make you review/recollect a few definitions/rules/facts/results related to convex functions:

- **Definition of a convex function:** A function f(x) is said to be convex if the following inequality holds true:

  $$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y) \quad \forall x, y \in Domain(f) \; and \; \alpha \in [0,1]$$

- **First-order condition of convexity:** A function f(x) which is differentiable is convex if the following inequality condition holds true:

  $$f(y) \;\ge\; f(x) + \nabla_x^T f(x)(y-x)\,; \quad \forall x, y \in Domain(f)$$

  Intuitively, this condition says that the tangent/first-order-taylor-series approximation of f(x) is globally an under-estimator.

- **Second-order condition of convexity:** A function f(x) which is twice-differentiable is convex if and only if its hessian matrix (matrix of second-order partial derivatives) is positive semi-definite, i.e.

  $$\forall z : \;\; z^T \, \nabla_x^2 f(x) \; z \ge 0 \;\; where \;\; \nabla_x^2 f(x) \;\; is \; the \; hessian$$

- **Sum/Linear-combination of two or more convex functions is also convex:** Let f(x) and g(x) be two convex functions. Then any linear combination of these two functions

  $$(\lambda_1 f + \lambda_2 g)(x) = \lambda_1 f(x) + \lambda_2 g(x)$$

  is also a convex function (this can be easily proved using the definition of the convex function).

Now notice that if we can prove that the two functions

$$-\log\left(h_\theta\left(x^i\right)\right) \;\; and \;\; -\log\left(1-h_\theta\left(x^i\right)\right)$$

are convex, then our objective function

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

must also be convex since any linear combination of two or more convex functions is also convex.

Let us now try to prove that

$$-\log\left(h_\theta(x)\right) = -\log\left(\frac{1}{1+e^{-\theta^T x}}\right) = \log\left(1+e^{-\theta^T x}\right)$$

About Me

My photo

Deepak Roy Chittajallu

Curiosity is the mother of all invention !!!

View my complete profile

Labels

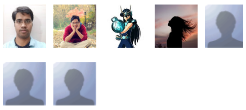
- Area of Circle (1)
- Arithmetic Progression (1)
- Circumference of a Circle (
- Convex functions (1)
- Convex Optimization (1)
- Convexity Conditions (1)
- Euler-poisson Integral (1)
- Fermat's Thoerem on Stationary Points (1)
- First-order condition of convexity (1)
- Gaussian Integral (1)
- Geometric Progression (1)
- Geometry (1)
- Logistic Regression (1)
- Machine Learning (1)
- Maxima and Minima (1)
- Multivariate Normal Distribution Integral (1)
- Normal Distribution Integrals (1)
- Principal Component Analysis (PCA) (1)
- Shortest Distance Curve Connecting Two Points (1)
- Univariate Normal Distrubution Integral (1)
- Variational Calculus (1)

Blog Archive

- ▼ 2011 (10)
  - ▼ October (2)
    - Proof for the first-order condition of convexit
    - Why is the error functio minimized in logistic re...
  - ► July (3)
  - ► June (5)

Followers

关注者（7 人）



关注

is a convex function of theta. In order to do this, we will use the second-order condition of convexity described above. Let us first compute the hessian matrix:

$$
\begin{aligned}
&grad: \\
\nabla_\theta \left[-\log\left(h_\theta(x)\right)\right] &= \nabla_\theta \left[\log\left(1+e^{-\theta^T x}\right)\right] \\
&= \left(\frac{-e^{-\theta^T x}}{1+e^{-\theta^T x}}\right) x \\
&= \left(\frac{1}{1+e^{-\theta^T x}} - 1\right) x \\
&= \left(h_\theta(x) - 1\right) x
\end{aligned}
$$

$$
\begin{aligned}
&hessian: \\
\nabla_\theta^2 \left[-\log\left(h_\theta(x)\right)\right] &= \nabla_\theta \left(\nabla_\theta \left[-\log\left(h_\theta(x)\right)\right]\right) \\
&= \nabla_\theta \left(\left(h_\theta(x) - 1\right) x\right) \\
&= h_\theta(x)\left(1 - h_\theta(x)\right) x x^T
\end{aligned}
$$

Now below is the proof that this hessian matrix is positive semi-definite:

$$
\begin{aligned}
\forall z: \quad z^T \nabla_x^2 \left(-\log\left(h_\theta(x)\right)\right) &= z^T \left[h_\theta(x)\left(1 - h_\theta(x)\right) x x^T\right] z \\
&= h_\theta(x)\left(1 - h_\theta(x)\right)\left(x^T z\right)^2 \geq 0 \qquad (2)
\end{aligned}
$$

Let us now try to prove that

$$
\begin{aligned}
-\log\left(1 - h_\theta(x)\right) &= -\log\left(1 - \frac{1}{1+e^{-\theta^T x}}\right) \\
&= -\log\left(\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right) \\
&= \theta^T x + \log\left(1 + e^{-\theta^T x}\right)
\end{aligned}
$$

is a convex function of theta. In order to do this, we will again use the second-order condition of convexity described above. Let us first compute its hessian matrix:

$$
\begin{aligned}
&grad: \\
\nabla_\theta \left[-\log\left(1 - h_\theta(x)\right)\right] &= \nabla_\theta \left[\theta^T x + \log\left(1 + e^{-\theta^T x}\right)\right] \\
&= x + \nabla_\theta \left[\log\left(1 + e^{-\theta^T x}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&hessian: \\
\nabla_\theta^2 \left[-\log\left(1 - h_\theta(x)\right)\right] &= \nabla_\theta \left(\nabla_\theta \left[-\log\left(1 - h_\theta(x)\right)\right]\right) \\
&= \nabla_\theta \left(x + \nabla_\theta \left[\log\left(1 + e^{-\theta^T x}\right)\right]\right) \\
&= \nabla_\theta^2 \left[-\log\left(h_\theta(x)\right)\right] \\
&\quad (we\ have\ proved\ in\ Eq.\ (2)\ above \\
&\quad that\ this\ is\ positive\ semi-definite)
\end{aligned}
$$

Above, we have proved that both

$$-\log\left(h_\theta\left(x^i\right)\right) \quad and \quad -\log\left(1 - h_\theta\left(x^i\right)\right)$$

are convex functions. And, the error/objective function of logistic regression

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

is essentially a linear-combination of several such convex functions. Now, since a linear combination of two or more convex functions is convex, we conclude that the objective function of logistic regression is convex.

Hence proved ...

Following the same line of approach/argument it can be easily proven that the objective function of logistic regression is convex even if regularization is used.

Posted by Deepak Roy Chittajallu at 9:38 PM

Labels: Logistic Regression, Machine Learning

12 comments:

Unknown November 2, 2015 at 3:42 PM

Just to correct a little mistake: only a positive linear combination of convex functions is guaranteed to be convex again. However, in the logistic regression case y_i are positive, so it works indeed.

Reply

Replies

for. October 17, 2020 at 8:54 AM

The development of artificial intelligence (AI) has propelled more programming architects, information scientists, and different experts to investigate the plausibility of a vocation in machine learning. Notwithstanding, a few newcomers will in general spotlight a lot on hypothesis and insufficient on commonsense application. machine learning projects for final year In case you will succeed, you have to begin building machine learning projects in the near future.

Projects assist you with improving your applied ML skills rapidly while allowing you to investigate an intriguing point. Furthermore, you can include projects into your portfolio, making it simpler to get a vocation, discover cool profession openings, and Final Year Project Centers in Chennai even arrange a more significant compensation.

Data analytics is the study of dissecting crude data so as to make decisions about that data. Data analytics advances and procedures are generally utilized in business ventures to empower associations to settle on progressively Python Training in Chennai educated business choices. In the present worldwide commercial center, it isn't sufficient to assemble data and do the math; you should realize how to apply that data to genuine situations such that will affect conduct. In the program you will initially gain proficiency with the specialized skills, including R and Python dialects most usually utilized in data analytics programming and usage; Python Training in Chennai at that point center around the commonsense application, in view of genuine business issues in a scope of industry segments, for example, wellbeing, promoting and account.

The Nodejs Projects Angular Training covers a wide range of topics including Components, Angular Directives, Angular Services, Pipes, security fundamentals, Routing, and Angular programmability. The new Angular TRaining will lay the foundation you need to specialise in Single Page Application developer. Angular Training

Reply

Unknown February 13, 2016 at 10:44 AM

This comment has been removed by the author.

Reply

Unknown February 13, 2016 at 11:09 AM

I think for equation(2) you lost a z in the left part.

Reply

Unknown August 14, 2017 at 11:42 PM

Thanks for post:
ship hỏa tốc sang Nepal
ship nhanh đi Nepal
ship nhanh tới Nepal
vận chuyển bưu phẩm đi Nepal
ship tốc độ đi Nepal
www.caycotacdunggi.info

Reply

Unknown March 9, 2018 at 9:26 AM

Awesome post, thank you.

Agree with above that (2) missing a z on L.H.S.

Reply

Effesian March 12, 2018 at 5:22 AM

I think in Eq.(2) you should write grad_{\theta} and not w.r.t. x.
This is of course just a typo.
Nice post btw.!

Reply

Unknown September 24, 2018 at 4:30 AM

I think things like this are really interesting. I absolutely love to find unique places like this. It really looks super creepy though!! Best Machine Learning institute in Chennai | Best Machine learning training | best machine learning training certification

Reply

Robotic Process Automation Tutorial November 4, 2018 at 9:49 PM

Very interesting blog.Thanks for sharing this much valuable information.Keep Rocking.
rpa training institute in chennai | rpa course fee in chennai | trending technologies list 2018

Reply

Diya shree July 27, 2019 at 2:46 AM

I feel very grateful that I read this. It is very helpful and very informative and I really learned a lot from it.

blue prism training in chennai | blue prism training in velachery | blue prism training and placement | best training institute for blue prism | blue prism course fee details | Best Blue Prism Training in Credo Systemz, Chennai | blue prism certification cost | blue prism certification training in chennai | blue prism developer certification cost

Reply

Javier T August 13, 2020 at 3:10 PM

Thank you a lot! It has been very useful to me :)

Reply

tuan March 23, 2021 at 8:48 PM

who know paper proves that this loss function is convex

Reply

Enter your comment...

Comment as: sm1les@qq.com

Sign out

Publish Preview

Notify me

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)