Why is the error function minimized in logistic regression convex?

3-3 minutes

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:

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

- **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.

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

- **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

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

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

Let us now try to prove that

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

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:

$$grad:$$

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

$$hessian:$$

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

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

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

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:

$grad:$

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

$hessian:$

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

Above, we have proved that both

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

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.