

Review of Machine Learning Assignment 1
(Ray) Shirui Lu
March 24, 2015

Convexity of Objective Functions in Logistic Regression

Background

Question: (ML part) Your results should be slightly different for every execution of the code, what is the source of the randomness?

Wrong Answer: w is initialized with random values, different sets of these values converges to different local minimal points.

Comments: First, the object function is a convex function, so there is no local minimal points, only one global minimal point.

Second, the source of randomness comes from random starting points, and insufficient iterations to finally reach the global minimal point, yet stops at different end point.

Furthermore, There is no need to reach the (theoretical) global minimal point: The improvement is trivial compared to the expensive computational cost, whereas may not obtain an optimal result (overfitting).

Brief Proof to the Convexity of Objective Function L^1

Proposition 1. Convex Functions of a Single Variable A twice-differentiable function f of a single variable x is convex if and only if

$$f''(x) \geq 0, \forall x \in \text{dom} f$$

We can extend the definition of convexity to higher dimension.

Definition 1. Second Order Condition for Convexity² A function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is twice differentiable. Then f is convex if and only if $\text{dom} f$ is a convex set, and its Hessian is positive semidefinite.³

Proof. For notational convenience, here we replace the original subscripts of i with k , since we will use i, j later to denote elements of Hessian.

$$\begin{aligned} L(w, w_0) &= - \sum_{k=1}^N t_k \log y_k + (1 - t_k) \log(1 - y_k) \\ &= \sum_{k=1}^N t_k \log(1 + e^{-w^T x_k - w_0}) + \sum_{k=1}^N (1 - t_k) (w^T x_k + w_0) + \sum_{k=1}^N (1 - t_k) \log(1 + e^{-w^T x_k - w_0}) \\ &= \sum_{k=1}^N \log(1 + e^{-w^T x_k - w_0}) + \sum_{k=1}^N (1 - t_k) (w^T x_k + w_0) \end{aligned}$$

Then we use a conclusion: **Linear combination of convex functions, with non-negative coefficients, is still a convex function**, the proof of this is trivial.

Hence The second term is linear combination of Affine Functions (in 2D its a line, 3D a plane), thus convex (all elements in its Hessian are 0).

For the first term, only need to examine $L_k = \log(1 + e^{-w^T x_k - w_0})$, if it's convex, L is linear combination of L_k , thus convex.

¹This a supplementary material, aiming at giving you a more thorough idea over the Logistic Regression loss function, not to be appeared in exams.

²For some other important concepts I mentioned in the tutorial, please refer to the B&V Book[1], which is available free online. (Convex Set[1, p. 23], Convex Function[1, p. 67]).

³Please refer to Sec **L is a convex function**

We then denote \mathbf{x}_k as \mathbf{u} , notice here \mathbf{x}_k is a vector (while \mathbf{x} is a matrix).

$$\begin{aligned}
L_k &= \log(1 + e^{-\mathbf{w}^T \mathbf{x}_k - w_0}) \\
\frac{\partial L}{\partial w_i} &= (y_k - 1)u_i, \\
\frac{\partial^2 L}{\partial w_i \partial w_j} &= \frac{\partial}{\partial w_j} \left(\frac{\partial L}{\partial w_i} \right) \\
&= \frac{\partial}{\partial w_j} (y_k - 1)u_i \\
&= y_k(1 - y_k)u_i u_j,
\end{aligned}$$

Hence the Hessian of L_k is:

$$\mathbf{H}_{L_k} = y_k(1 - y_k)\mathbf{u}\mathbf{u}^T$$

Then we show the Hessian of L_k is **positive semidefinite**:

$$\begin{aligned}
\forall \mathbf{z} : \mathbf{z}^T \mathbf{H}_{L_k} \mathbf{z} &= \mathbf{z}^T [y_k(1 - y_k)\mathbf{u}\mathbf{u}^T] \mathbf{z} \\
&= y_k(1 - y_k)(\mathbf{u}^T \mathbf{z})^2, \text{ notice that } y_k \in (0, 1) \\
&\geq 0
\end{aligned}$$

\mathbf{H}_{L_k} is positive semidefinite, thus L is a convex function. □

References

- [1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.