hw1

wangxinshi47

February 2023

1 Introduction

$f(\omega') = \frac{1}{n}\sum_{i=1}^{n}\log(1 + exp(-y_i(\omega_0 + \omega^T x_i))) + \lambda||\omega||_2^2$, where $\omega' = [\omega_0; \omega]$ is the vertical concatenation of the bias and the feature coefficients.

The gradient of the loss function with respect to ω' can be computed as:

$$\nabla f(\omega') = \begin{bmatrix} \frac{\partial f(\omega')}{\partial \omega_0} \\ \frac{\partial f(\omega')}{\partial \omega} \end{bmatrix}$$

Where $\frac{\partial f(\omega')}{\partial \omega_0}$ and $\frac{\partial f(\omega')}{\partial \omega}$ can be computed as follows:

$$\frac{\partial f(\omega')}{\partial \omega_0} = \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i \exp(-y_i(\omega_0+\omega^T(x_i)))}{1+exp(-y_i(\omega_0+\omega^T x_i))} = \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i}{1+\exp(y_i(\omega_0+\omega^T x_i))}$$

$$\frac{\partial f(\omega')}{\partial \omega} = \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i x_i \exp(-y_i(\omega_0+\omega^T x_i))}{1+exp(-y_i(\omega_0+\omega^T x_i))} = \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i x_i}{1+\exp(y_i(\omega_0+\omega^T x_i))} + 2\lambda\omega$$

Therefore, the gradient of the loss function with respect to ω' is:

$$\nabla f(\omega') = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i}{1+\exp(y_i(\omega_0+\omega^T x_i))} \\ \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i x_i}{1+\exp(y_i(\omega_0+\omega^T x_i))} + 2\lambda\omega \end{bmatrix}$$

To compute the Hessian matrix, we need to find the following partial derivatives:

$$\nabla^2 f(\omega') = \begin{bmatrix} \frac{\partial^2 f(\omega')}{\partial \omega_0^2} & \frac{\partial^2 f(\omega')}{\partial \omega_0 \partial \omega} \\ \frac{\partial^2 f(\omega')}{\partial \omega \partial \omega_0} & \frac{\partial^2 f(\omega')}{\partial \omega^2} \end{bmatrix}$$

Where $\frac{\partial^2 f(\omega')}{\partial \omega_0^2}$, $\frac{\partial^2 f(\omega')}{\partial \omega_0 \partial \omega}$, $\frac{\partial^2 f(\omega')}{\partial \omega \partial \omega_0}$, and $\frac{\partial^2 f(\omega')}{\partial \omega^2}$ can be computed as follows:

$$\frac{\partial^2 f(\omega')}{\partial \omega_0^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}$$

$$\frac{\partial^2 f(\omega')}{\partial \omega_0 \partial \omega} = \frac{\partial^2 f(\omega')}{\partial \omega \partial \omega_0} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i x_i \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}$$

$$\frac{\partial^2 f(\omega')}{\partial \omega^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 x_i x_i^T \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2} + 2\lambda$$

Therefore, the Hessian matrix of the loss function with respect to ω' is (There is a transpose on x in the second element because the hessian matrix is a square matrix):

$$\nabla^2 f(\omega') = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2} & \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 x_i^T \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2} \\ \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 x_i \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2} & \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2 x_i x_i^T \exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2} + 2\lambda \end{bmatrix}$$

To argue the logistic regression optimization problem is a convex optimization problem, we can rewrite the hessian matrix as $\frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2\exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}\begin{bmatrix}1 & x_i^T\\ x_i & x_i x_i^T\end{bmatrix}+$ $2\lambda\begin{bmatrix}0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1\end{bmatrix}$. Let $A=\begin{bmatrix}1\\ x_i\end{bmatrix}$, we can easily decompose $\begin{bmatrix}1 & x_i^T\\ x_i & x_i x_i^T\end{bmatrix}$ since $AA^T=\begin{bmatrix}1\\ x_i\end{bmatrix}\begin{bmatrix}1 & x_i^T\end{bmatrix}$. By multiplying A with the square root of $\frac{1}{n}\frac{y_i^2\exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}$, we know $\frac{1}{n}\frac{y_i^2\exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}\begin{bmatrix}1 & x_i^T\\ x_i & x_i x_i^T\end{bmatrix}$ is positive semi-definite. Since the sum of PSD matrix is PSD, we have $\frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2\exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}\begin{bmatrix}1 & x_i^T\\ x_i & x_i x_i^T\end{bmatrix}$ as a PSD matrix. Since $2\lambda\begin{bmatrix}0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1\end{bmatrix}$ is a diagonal matrix and has all of its values positive, it is also a positive semi-definite matrix. Since the sum of positive semi-definite matrices are positive semi-definte, $\frac{1}{n}\sum_{i=1}^{n}\frac{y_i^2\exp(y_i(\omega_0+\omega^T x_i))}{(1+\exp(y_i(\omega_0+\omega^T x_i)))^2}\begin{bmatrix}1 & x_i^T\\ x_i & x_i x_i^T\end{bmatrix}+$ $2\lambda\begin{bmatrix}0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1\end{bmatrix}$ is a positive semi-definite matrix. Since the hessian is PSD, we can conclude the logistic regression optimization problem is convex.