Homework 2

NE 795-001: Fall 2023

Arjun Earthperson

October 29th, 2023

Problem 1 - Logistic Regression

In the given equation:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m}[(y^{(i)} - \sigma(\beta^T x^{(i)})] \cdot x_j^{(i)}$$

- $LL(\beta)$ is the log-likelihood function, which measures the goodness of fit of a logistic regression model. The goal is to find the parameters β that maximize this function.
- β is a vector of parameters in the logistic regression model, and β_j is the j-th parameter.

The log-likelihood function $LL(\beta)$ for a logistic regression model is given by:

$$LL(\beta) = \sum_{i=1}^{m}[y^{(i)} \log(\sigma(\beta^T x^{(i)})) + (1 - y^{(i)}) \log(1 - \sigma(\beta^T x^{(i)}))]$$

where $y^{(i)}$ is the observed output for the i-th sample, $\sigma(\beta^T x^{(i)})$ is the predicted probability for the i-th sample, and $\sigma(\beta^T x^{(i)})$ is given by the logistic function:

$$\sigma(\beta^T x^{(i)}) = \frac{1}{1 + e^{-\beta^T x^{(i)}}}$$

The derivative of the log-likelihood function $LL(\beta)$ with respect to the parameter β_j can be computed using the chain rule:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \frac{\partial LL(\beta)}{\partial \sigma(\beta^T x^{(i)})} \cdot \frac{\partial \sigma(\beta^T x^{(i)})}{\partial \beta^T x^{(i)}} \cdot \frac{\partial \beta^T x^{(i)}}{\partial \beta_j}$$

The derivative of the log-likelihood function with respect to $\sigma(\beta^T x^{(i)})$ is:

$$\frac{\partial LL(\beta)}{\partial \sigma(\beta^T x^{(i)})} = \frac{y^{(i)}}{\sigma(\beta^T x^{(i)})} - \frac{1 - y^{(i)}}{1 - \sigma(\beta^T x^{(i)})}$$

The derivative of $\sigma(\beta^T x^{(i)})$ with respect to $\beta^T x^{(i)}$ is:

$$\frac{\partial \sigma(\beta^T x^{(i)})}{\partial \beta^T x^{(i)}} = \sigma(\beta^T x^{(i)})(1 - \sigma(\beta^T x^{(i)}))$$

The derivative of $\beta^T x^{(i)}$ with respect to β_j is simply $x_j^{(i)}$.

Therefore, the derivative of the log-likelihood function with respect to β_j is:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \left(\frac{y^{(i)}}{\sigma(\beta^T x^{(i)})} - \frac{1 - y^{(i)}}{1 - \sigma(\beta^T x^{(i)})} \right) \cdot \sigma(\beta^T x^{(i)})(1 - \sigma(\beta^T x^{(i)})) \cdot x_j^{(i)}$$

We can distribute the terms $\sigma(\beta^T x^{(i)})$ and $(1 - \sigma(\beta^T x^{(i)}))$ inside the parentheses:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \left(y^{(i)}(1 - \sigma(\beta^T x^{(i)})) - (1 - y^{(i)})\sigma(\beta^T x^{(i)}) \right) \cdot x_j^{(i)}$$

Then, we can distribute $y^{(i)}$ and $(1 - y^{(i)})$ inside the parentheses:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \left(y^{(i)} - y^{(i)}\sigma(\beta^T x^{(i)}) - \sigma(\beta^T x^{(i)}) + y^{(i)}\sigma(\beta^T x^{(i)}) \right) \cdot x_j^{(i)}$$

Notice that the terms $-y^{(i)}\sigma(\beta^T x^{(i)})$ and $y^{(i)}\sigma(\beta^T x^{(i)})$ cancel out:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \left(y^{(i)} - \sigma(\beta^T x^{(i)}) \right) \cdot x_j^{(i)}$$

So, the derivative of the log-likelihood function with respect to β_j simplifies to:

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} (y^{(i)} - \sigma(\beta^T x^{(i)})) \cdot x_j^{(i)}$$

where

- m is the number of observations in the dataset.
- $y^{(i)}$ is the actual class label of the i-th observation.
- $\sigma(\beta^T x^{(i)})$ is the predicted probability that the i-th observation belongs to the positive class, according to the logistic regression model. Here, σ is the logistic sigmoid function, β^T is the transpose of the parameter vector, and $x^{(i)}$ is the feature vector of the i-th observation.
- $x_j^{(i)}$ is the j-th feature of the i-th observation.