Gradient Descent and Regularization in Logistic Regression

Tran Quoc Thai
ID student: 2370759

April 9, 2025

1 Gradient Descent for Logistic Regression

In logistic regression, the cost function is defined as:

$$J(\mathbf{w}) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log h_{\mathbf{w}}(x^{(i)}) + (1-y^{(i)})\log(1-h_{\mathbf{w}}(x^{(i)}))\right]. \tag{1}$$

The **gradient of the cost function** with respect to each weight w_j is given by:

$$\frac{\partial J}{\partial w_j} = \frac{1}{m}\sum_{i=1}^{m}(h_{\mathbf{w}}(x^{(i)}) - y^{(i)})x_j^{(i)}. \tag{2}$$

1.1 Computing the Gradient for One Iteration

From **Question 7**, the logistic regression prediction for T_1 was:

$$h_{\mathbf{w}}(x_1) = 0.9989. \tag{3}$$

Assuming T_1's actual **RiskLevel** is $y_1 = 0$ (Low Risk), the gradient components are computed as:

$$\frac{\partial J}{\partial w_0} = (0.9989 - 0) \times 1 = 0.9989,$$
$$\frac{\partial J}{\partial w_1} = (0.9989 - 0) \times 37 = 36.9593,$$
$$\frac{\partial J}{\partial w_2} = (0.9989 - 0) \times 705 = 703.7895.$$

1.2 Gradient Descent Update Step

Using **learning rate** $\alpha = 0.01$, the updated parameters are:

$$
\begin{aligned}
w_0 &:= w_0 - \alpha \frac{\partial J}{\partial w_0} = 0.5 - (0.01 \times 0.9989) = 0.4900, \\
w_1 &:= w_1 - \alpha \frac{\partial J}{\partial w_1} = -0.02 - (0.01 \times 36.9593) = -0.3896, \\
w_2 &:= w_2 - \alpha \frac{\partial J}{\partial w_2} = 0.01 - (0.01 \times 703.7895) = -7.0279.
\end{aligned}
$$

2 Impact of Regularization

Regularization helps prevent overfitting by adding a **penalty term** to the cost function. The two common types are:

- **L1 Regularization (Lasso)**: Adds $\lambda \sum |w_j|$, encouraging sparsity.
- **L2 Regularization (Ridge)**: Adds $\lambda \sum w_j^2$, preventing large weights.

With **L2 Regularization**, the new gradient update rules become:

$$
\frac{\partial J}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (h_{\mathbf{w}}(x^{(i)}) - y^{(i)}) x_j^{(i)} + \lambda w_j. \tag{4}
$$

Applying **regularization with** $\lambda = 0.1$:

$$
\begin{aligned}
\frac{\partial J}{\partial w_0} &= 0.9989, \\
\frac{\partial J}{\partial w_1} &= 36.9593 + 0.1(-0.02) = 36.9573, \\
\frac{\partial J}{\partial w_2} &= 703.7895 + 0.1(0.01) = 703.7905.
\end{aligned}
$$

This ensures weights are **controlled**, preventing excessive updates.

3 Why Regularization is Necessary for This Dataset

- **Large Feature Values**: Since **Age** and **CreditScore** are large, their associated weights could grow significantly, leading to unstable models.
- **Prevents Overfitting**: Regularization helps maintain **generalizability** by discouraging extreme values.
- **Improves Numerical Stability**: Reducing large updates helps convergence.

4 Conclusion

In conclusion, the application of gradient descent in logistic regression provides a systematic approach for optimizing model parameters by iteratively reducing the cost function. Without regularization, gradient updates can lead to large parameter values, resulting in overfitting and decreased predictive performance on unseen data. The introduction of $L2$ regularization serves as an effective mechanism to control parameter growth, prevent overfitting, and improve the robustness of the model. By incorporating regularization terms into the optimization process, the model achieves better stability, generalization capability, and interpretability, ultimately leading to more reliable predictions in practical applications.