Logistic Regression Model

Logistic regression is a statistical model used for binary classification tasks, which predicts the probability that an observation belongs to one of two classes. It is particularly useful when the response variable is categorical and dichotomous.

Hypothesis Function

The hypothesis in logistic regression predicts the probability that a given observation falls into the class labeled as "1". This probability is modeled using the logistic function or sigmoid function:

$$h_\theta(x) = \sigma(\theta^T x)$$
$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Here:

- θ is the parameter vector, including the intercept and coefficients.
- x is the feature vector, including a constant term 1 to accommodate the intercept.
- $\sigma(z)$ is the logistic or sigmoid function.

Decision Boundary

The decision boundary, which separates the classes, is defined by the set of points where $h_\theta(x) = 0.5$. This occurs when:

$$\theta^T x = 0$$

Cost Function

The cost function in logistic regression, also known as the log loss, measures the performance of a classification model where the prediction input is a probability value between 0 and 1. The cost function is defined as:

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

Where:

- m is the number of training examples.
- $y^{(i)}$ is the actual label of the ith training example.
- $h_\theta(x^{(i)})$ is the predicted probability that the ith training example belongs to class 1.

Parameter Estimation: Gradient Descent

To find the parameters θ that minimize the cost function, we can use gradient descent. The gradient of the cost function with respect to θ is:

$$\nabla J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$$

The update rule in gradient descent is:

$$\theta := \theta - \alpha \nabla J(\theta)$$

Where α is the learning rate.

Convergence

The algorithm iteratively adjusts the parameters to minimize the cost function until it converges to a minimum, ideally representing the best parameter values for predicting the target variable.

Exam Problem: Logistic Regression for University Admissions

Context:

A university admissions committee wants to predict which applicants will accept an offer of admission based on two features: undergraduate GPA (Grade Point Average) and a personal interview score. Historical data from previous admissions cycles is available, including GPA, interview score, and whether the applicant accepted the offer (1 for accept, 0 for decline).

Given Data Sample:

Applicant	GPA	Interview Score	Accepted?
1	3.5	8	1
2	3.7	7	1
3	2.8	6	0
4	3.0	5	0

Objective:

Use logistic regression to model the probability that an applicant will accept the offer based on their GPA and interview score.

Tasks:

1. **Model Formulation**:

- Define the logistic regression model that relates the probability of acceptance to GPA and interview score. Include the hypothesis function using the logistic function.

2. **Parameter Estimation**:

   - Assume initial values of parameters θ=[0,0,0] (intercept, GPA coefficient, interview score coefficient). Calculate the predicted probability for each applicant using these initial parameters.

3. **Cost Calculation**:

   - Using the logistic regression cost function, calculate the cost (log loss) for the given initial parameters based on the sample data.

4. **Gradient Calculation**:

   - Compute the gradient of the cost function with respect to each parameter using the initial parameter values.

5. **Interpretation and Discussion**:

   - Discuss how changes in GPA and interview scores might affect the probability of an applicant accepting the offer. Consider explaining what the coefficients of the logistic regression model tell us in practical terms.

Solution

1. Model Formulation:

The logistic regression model can be defined as:

$P(Y = 1|x) = \sigma(\theta_0 + \theta_1 \times \text{GPA} + \theta_2 \times \text{Interview Score})$

Where $\sigma(z)$ is the logistic function:

$\sigma(z) = \frac{1}{1+e^{-z}}$

2. Parameter Estimation:

Given initial parameter values $\theta_0 = 0, \theta_1 = 0, \theta_2 = 0$:

$z = \theta_0 + \theta_1 \times \text{GPA} + \theta_2 \times \text{Interview Score} = 0$

$\sigma(z) = \frac{1}{1+e^{-0}} = 0.5$

This implies that the predicted probability of accepting the offer, regardless of GPA and Interview Score, is 0.5 for all applicants.

3. Cost Calculation:

Using the logistic regression cost function:

$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [y_i \log(\sigma(z_i)) + (1 - y_i) \log(1 - \sigma(z_i))]$

Given $m = 4$ (four applicants), $\sigma(z_i) = 0.5$ for all i:

$J(\theta) = -\frac{1}{4}[(1 \cdot \log(0.5) + (1-1) \cdot \log(1-0.5)) + (1 \cdot \log(0.5) + (1-1) \cdot \log(1-0.5)) + (0 \cdot \log(0.5) + (1-0) \cdot \log(1-0.5)) + (0 \cdot \log(0.5) + (1-0) \cdot \log(1-0.5))]$

$J(\theta) = -\frac{1}{4}[(-0.6931) + (-0.6931) + (-0.6931) + (-0.6931)]$

$J(\theta) = -\frac{1}{4}[-2.7724] = 0.6931$

4. Gradient Calculation:

The gradient of the cost function $\nabla J(\theta)$ for each parameter is:

$\frac{\partial J}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (\sigma(z_i) - y_i) x_{ij}$

$\text{For } \theta_0: \frac{\partial J}{\partial \theta_0} = \frac{1}{4}[(0.5-1) \times 1 + (0.5-1) \times 1 + (0.5-0) \times 1 + (0.5-0) \times 1] = \frac{1}{4}[-0.5 - 0.5 + 0.5 + 0.5] = 0$

$\text{For } \theta_1 \text{ and } \theta_2, \text{ similar calculations apply, resulting in 0 due to the simplification and init 0.}$

5. Interpretation and Discussion:

- **Effect of GPA and Interview Scores**: With non-zero coefficients, an increase in GPA and Interview Scores would increase the logistic function input, pushing the output closer to 1, hence increasing the probability of acceptance. The positive coefficients indicate direct positive influence on the acceptance probability.
- **Coefficient Interpretation**: In practical terms, a positive coefficient for GPA or Interview Score would suggest that higher values of these predictors are associated with a higher likelihood of acceptance. Conversely, a negative coefficient would suggest an inverse relationship.