Om Shete

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Assignment 4

Q.1 Explain the intuition behind Logistic Regression in detail. Is the decision boundary Linear or Non-linear in the case of a Logistic Regression model? Also explain the impact of outliers on logistic regression.

⇒

- **Intuition behind logistic Regression:**

- Logistic regression is supervised learning algorithm used primarily for binary classification problems.
- Despite its name, logistic regression is actually a classification algorithm, not a regression one
- The regression term is used because the model estimates probabilities and probabilities are continuous values between 0 and 1.
- The core idea behind logistic regression is that we want to model the probability that a given input belongs to a particular class.
- This probability is modeled as sigmoid (logistic) function applied to a linear combination of input features.

- **Is the Decision Boundry Linear or Non-Linear in Logistic Regression?**

- In ~~linear~~ regression logistic regression, the decision boundry is linear.
- This is because the model is fundamentally

based on a linear combination of input features.

- The sigmoid function transforms the linear combination into probabilities, but the decision boundary remains linear in the feature space.
- For e.g., for two feature x_1 and x_2, the decision boundary can be expressed as a line:

$$w_1 x_1 + w_2 x_2 + b = 0$$

- If you want non-linear decision boundaries, you can extend logistic regression by including polynomial feature or other types of feature transformation.
- In that case, the decision boundary can become non-linear in the original feature space.
- However, the core logistic regression model is always linear in terms of its input feature.

• <u>Impact of outliers on Logistic Regression:</u>

- Outliers are the data points that are significantly different from the majority of the data.
- In logistic regression, outliers can have a significant impact because logistic regression is a parametric model.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Q.2 What are the assumptions made in logistic regression? Can we solve the multiclass classification problem using logistic regression? If yes then how?

=)

- Assumption Made in Logistic Regression:

1) The observations are independent
=> Logistic regression assumes that the observations in the dataset are independent of each other. That is, the observations should not come from repeated measurements of the same individual or be related to each other in any way.

2) There is no multicollinearity among explanatory variables.
=> Multicollinearity occurs when two or more explanatory variables are highly correlated to each other, such that they do not provide unique or independent information in regression model.

3) The sample size is sufficiently large
=> Logistic regression assumes that the sample size of the dataset is large enough to draw valid conclusions from the fitted logistic regression model.

Solving Multiclass classification problem:

- Yes, logistic regression can be extended to solve multiclass classification problems.
- This extension is done using multinominal logistic regression or by using one-vs-rest (OvR) and one-vs-one (ovo) stratergies.
- Multinominal logistic regression is the direct extension of logistic regression for multiclass classification problems where the outcomes variable can have three or more categories.
- It uses softmax function to generalize logistic regression for multiple classes.
- The softmax function computes probabilities for each class label, and the label with the highest probability is chosen as the predicted class.
- In the One-vs-Rest approach, we fit one binary logistic regression classifier per class.
- For each class, we treat that class as the positive class (1) and all the other classes as the negative class (0).
- This results in multiple binary classifier.
- When predicting, we compute the probability that the sample belongs to each class and choose the class with the highest probability.

- In the one-vs-one approach, we build a binary classifier for every pair of classes.
- For each pair of classes, we train a classifier to distinguish between those two classes only.
- For a problem with K classes, we create $\frac{K(K-1)}{2}$ classifiers.
- For prediction, we apply each classifier and choose the class that wins the most pairwise comparisons.

Q.3 Why is Logistic regression termed as Regression and not as classification?

⇒

- Logistic regression is called "regression" even though it is used for classification because of its origins and nature of the model.
- In linear regression, we predict a continous outcome using a linear combination of the input feature:

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

- Logistic regression also computes a linear combination of input features, but it applies the logistic function to transform the output into a probability between 0 and 1.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

$$P(y=1|x) = \frac{1}{1+e^{-(w_1x_1+w_2x_2+\dots+w_nx_n+b)}}$$

- The logistic function converts the continuous output of the linear combination into a probability, allowing it to be used for classification.
- The term "classification" refers to the final task of assigning labels (0 and 1) but logistic regression describe the method of calculating the probability of those models.
- Logistic regression doesn't directly classify; instead, it regresses on the features to estimate probabilities, and classification is the step that occurs based on the estimated probabilities.
- Although logistic regression is used for classification, it fits a continous decision boundary
- The outcome of the linear combination is continuous, which is then mapped to a range of 0 to 1 using the sigmoid function.
- This continuous output is a hallmark of regression model.

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Q.4 Can we use Mean square Error as a cost function for logistic regression? Justify your answer.

⇒

- No, Mean squared Error (MSE) is not appropriate as a cost function for logistic regression.

1) MSE assumes a linear relationship

⇒ MSE is primarily used in linear regression, where the relationship between the dependent and independent variables is assumed to be linear.

- It works well in this context because the predicted value and true values are both continuous, and MSE measures the squared difference between them.

2) Non-linearity of the Sigmoid Function

⇒ Logistic regression uses the sigmoid function to map the linear combination of the input features to a probability value.

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

- The sigmoid function is non-linear and bounded between 0 and 1, where MSE assumes the output can take on any real value.

THADOMAL SHAHANI TSEC ENGINEERING COLLEGE

3) MSE leads to Non-Convex cost function

⇒ When using the sigmoid function in logistic regression, the shape of the cost function derived from MSE often become non-convex.

- A non-convex cost function can make it difficult for optimization algorithms like gradient descent to find the global minimum efficiently.

4) Inappropriate Gradient with MSE

⇒ The derivate of MSE is linear, which may not be suitable for updating the weights in the logistic regression since the output is not linear due to the sigmoid transformation.

- With log-loss, the gradient of the cost function is better aligned with logistic function, resulting in more effective weights updates during optimization.

Q.5 Compare Naive Bayesian with logistic regression classifier.

=>

Naive Bayes	logistic Regression
1) Model type is generative.	1) Model type is discriminative.
2) Assumes features are independent of each other.	2) No assumption about independence.
3) Model complexity is relatively simple, computationally efficient	3) More complex, allowing for modeling intricate relationships.
4) well-suited for categorical data, effective for discrete features and text classification.	4) Versatile, can handle both numerical and categorical features.
5) Highly interpretable due to simplicity and the assumption of feature independence.	5) Good interpretability through coefficients, indicating the strengths and direction of feature relationships.
6) Performs well with small dataset.	6) May require a larger dataset to avoid overfitting.
7) suitable for text classification, spam filtering, etc.	7) suitable for customer churn prediction, credit scoring.

27/9/24