Binary and Multinomial Logistic Regression Models

 Lisa Hall

University of Cumberlands

MSDS-530-A01 - First Bi-term

Dr. Afshin Zarenejad

February 19, 2025

**Binary and Multinomial Logistic Regression Models**

Despite being valuable and easy to apply, logistic regression models are not used as often as they could be in data analysis (Fávero & Belfiore, 2018). Logistic regression models are like linear regression models, except outcomes are binary, such as predicting pass/fail, yes/no, or 0/1. Multinomial logistic regression models can estimate the probability of more than two values. For example, predicting a customer’s preferred dish, like eggs, bacon, or toast, would require multinomial logistic regression.

Logistic regression models differ from linear regression models in output type. Linear regression is continuous, like predicting pricing. Logistic regression predicts probabilities related to a specific outcome (Bruce & Bruce, 2019). Both models estimate relationships between independent variables and an expected result; however, the applications can be very different. Linear regression could forecast sales revenue based on marketing, whereas logistic regression may better predict whether a student will pass an exam.

**Logistic Regression Example**

Python libraries like sklearn.linear\_model, sklearn.model\_selection, sklearn.Metrics allow logistic regression models to be performed quickly and accurately(Scikit-learn developers, n.d.). Sklearn. Datasets allow for the quick selection of example datasets that can be used as practice sets to hone the skills needed to master the creation of logistic models. Figure 1 is an example of Python code needed to create a logistic regression model. The dataset used is a sample breast cancer dataset. X contains features independent variables like mean radius, mean texture, and mean perimeter, and y contains target variables like malignant (0) and benign (1). X is the input, and y is the output of what the model is trying to predict.

**Figure 1**

**Logistic Regression Python code**

A screen shot of a computer program

AI-generated content may be incorrect.

Further exploring Figure 1, the logistic regression() function. We set the max iteration of the model to 1000; the more iterations, the better chance the model finds the best parameters. The model is trained using the model.fit() function; it learns the relationship between the features (x\_train) and the target (y\_train). We can then make predictions using the model.predict(X-test); each prediction would be either 0 or 1, malignant or benign. Print(classification report(y\_test, y\_pred)) generates metrics that help evaluate the model’s performance. In Figure 2, we can view the output. Precision indicates how many predicted outcomes were correct; when the model predicted malignant, it was correct 89% of the time; for positive, it was correct 96% of the time. The F1-Score indicates how well the model performed; for malignant, it indicates good performance; and for benign, it indicates excellent performance. Support indicates that 52 malignant and 108 benign cases were in the test set. The weighted average indicates that the mode’s overall performance was 94%.

**Figure 2**

**Regression Model Output**

A screenshot of a computer

AI-generated content may be incorrect.

**Real World Applications**

Logistic regression models are beneficial when predicting healthcare outcomes, as noted in the above example. Further data could help predict survival rates of malignancies with or without treatment. Another use could be predicting whether an email is or is not spam (IBM, n.d.). Multinomial logistic regression could be applied to prediction preferences when there are more than two choices, like movie preferences. A wide range of possibilities exist for using logistic regression when predicting possible outcomes.

**Conclusion**

Classification tools like binary and multinomial logistic regression are important models across domains. Decision makers rely on model accuracy for informed decision makers. Tools like Python allow for quick and accurate data processing in logistic regression models. Real-world applications include making predictions that could impact disease diagnosis and spam prediction. The practical use of logistic regression makes it a helpful choice in data science.

**References**

Bruce, P., & Bruce, A. (2019). *Practical statistics for data scientists: 50+ essential concepts using R and Python* (2nd ed.). O'Reilly Media.

Fávero, L. P., & Belfiore, P. (2018). *Data Science for Business and Decision Making* (1st ed.). Academic Press. https://doi.org/10.1016/C2016-0-01101-4

IBM. (n.d.). *What is logistic regression?* IBM. Retrieved February 20, 2025, from https://www.ibm.com/think/topics/logistic-regression

Scikit-learn developers. (n.d.). *sklearn.linear\_model.LogisticRegression*. Scikit-learn. Retrieved [February 20, 2025], from https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LogisticRegression.html

Scikit-learn developers. (n.d.). *sklearn.datasets.load\_breast\_cancer*. Scikit-learn. Retrieved [February 20, 2025], from https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_breast\_cancer.html