**Regression**

Alissa Owens

11/8/2023

CS 430

**Introduction:**

In this discussion, we delve into two fundamental statistical techniques: linear regression and logistic regression. Linear regression serves as a powerful tool for predicting continuous numeric values, while logistic regression specializes in binary classification, particularly in determining the probability of an observation belonging to one of two classes. We also touch upon essential evaluation metrics, including R-squared (R^2) and Root Mean Squared Error (RMSE), as they relate to linear regression models. Additionally, we explore the renowned Iris dataset, a staple in the fields of statistics and machine learning. This dataset, featuring three distinct species of iris flowers and their associated attributes, is frequently employed to illustrate classification tasks, where the primary objective is to predict the species of iris plants based on these attributes. The choice of regression technique and evaluation metrics is pivotal and should align with the specific nature of the problem being addressed.

**Explanation of the Models Used:**

What is Linear Regression? Linear regression is a way to figure out how one thing (like the price of a house) is related to another thing (like the size of the house). It helps you find a straight line that best fits the data and allows you to make predictions or understand how changes in one thing affect the other. For example, if you know the size of a house, linear regression can help you predict its price based on a historical pattern. It's like drawing a line through a scatterplot of data points to find a pattern or relationship.

What is Logistic regression? Logistic regression is a statistical method used for binary classification. It predicts the probability of an observation belonging to one of two classes (0 or 1). It uses a logistic (sigmoid) function to model these probabilities, and the model learns coefficients that determine the influence of input features on the probability of the positive class. Logistic regression is widely used for tasks like spam detection and medical diagnosis in various fields.

Linear regression predicts continuous numeric values, while logistic regression predicts binary outcomes. Linear regression uses a linear equation, while logistic regression uses the logistic function for modeling. Linear regression is for continuous variables, while logistic regression is for binary classification. Both learn coefficients from data and are used in supervised learning.

R-squared (R^2) measures the proportion of the variance in the dependent variable (target) that is explained by the independent variables in the model. R^2 values range from 0 to 1, with higher values indicating a better fit. In the context of linear regression, R^2 helps evaluate the model's ability to explain the variance in the continuous target variable.

Root Mean Squared Error (RMSE) measures the average magnitude of the residuals (differences between predicted and actual values) in the linear regression model. Smaller RMSE values indicate better model accuracy in predicting continuous values.

**Discussion of Dataset:**

the Iris dataset, which is a well-known dataset in the field of statistics and machine learning. It contains information about three different species of iris flowers: Iris-setosa, Iris-versicolor, and Iris-virginica. The dataset consists of 150 instances, with each class having 50 instances.

The dataset includes five attributes for each instance, which are:

Sepal Length (in centimeters)

Sepal Width (in centimeters)

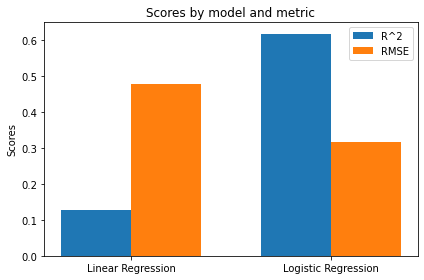
Petal Length (in centimeters)

Petal Width (in centimeters)

Class (the species of iris plant)

The goal of this dataset is typically to predict the class (species) of the iris plant based on the provided attribute values. It's a common dataset used for practicing and demonstrating classification techniques in machine learning.

**Results:**



**Conclusion:**

In conclusion, the information provided highlights the fundamental differences between linear regression and logistic regression, where the former is used for predicting continuous values and the latter for binary classification. Key evaluation metrics, such as R-squared and RMSE, are discussed in the context of linear regression. The dataset in question, the Iris dataset, is introduced as a widely recognized dataset for classification tasks involving three species of iris flowers. Its primary purpose is to predict the species of iris plants based on their attributes, making it a common benchmark in the field of machine learning. It's essential to choose the appropriate regression technique and evaluation metrics according to the specific problem at hand.