

## What's the difference between polynomial/linear regression and logistic regression?

The fundamental difference lies in their purpose: linear and polynomial regression are used for **regression** tasks, which predict a continuous numerical value. Logistic regression is for **classification**, predicting a categorical outcome.

**Output Type:** Regression models predict a **continuous numerical value** that can fall anywhere within a range, such as a house price or a person's age. In contrast, logistic regression outputs a **probability** between 0 and 1, which is then used to classify data into discrete categories, like "fraud" or "not fraud."

**Model Goal:** The goal of linear and polynomial regression is to find the best-fit line or curve that minimizes the **sum of squared errors** between the model's predictions and the actual data points. Logistic regression's goal is to find a line or curve that best separates the data points into their respective classes.

In the fraud detection code, which is a **classification** problem, both linear and polynomial regression were used in a non-standard way.

(1) Linear and polynomial models produce a continuous output. To classify these outputs as "fraud" or "not fraud," a manual **threshold** (in this case, 0.5) was applied.

(2) Logistic regression, however, is the proper model for this type of problem. It directly outputs the **probability** of a transaction being fraudulent, making the classification at a 0.5 threshold a natural part of the model's design.

Although all three models performed well, logistic regression is the most appropriate and robust choice because it is built to provide a probabilistic outcome, which is the direct goal of a classification task.