What's the difference between polynomial/linear regression and logistic regression? Explain when you would use each one and how their outputs are different.

The main difference between linear/polynomial regression and logistic regression lies in the type of problem they solve. Linear and polynomial regression are used for regression problems, and their goal is to predict a continuous numerical output. Meanwhile, logistic regression is used for classification problems, where it is used to predict a discrete categorical output.

Linear Regression models the relationship between features (X) and the target variable (y) as a straight line. It assumes a linear relationship between the variables. Polynomial Regression is an extension that models the relationship as a curve using a polynomial function. This is useful when the relationship between the features and the target isn't linear. I would use these methods to predict things like housing prices, a person's age, or the temperature. The output is a continuous numerical value. Performance of these models is typically measured using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R squared score, which quantify the difference between the predicted and actual continuous values.

Logistic Regression is used for predicting categorical values (classifying data). It's like linear regression in that it calculates a weighted sum of the input features. However, instead of using this value directly, it passes it through a sigmoid function. The sigmoid function transforms the output of the linear equation into a value between 0 and 1. This value represents the probability of the input belonging to a specific class. I would use logistic regression to predict whether an email is spam or not, if a customer will churn, or if a tumor is malignant or benign. The output is a probability, which is then used to assign the data point to a discrete. Performance is measured using metrics suited for classification problems, such as a confusion matrix, precision, recall, and the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score. These metrics assess the model's ability to correctly classify instances into their respective categories.