Answer 1

Logistic regression aims to model the probability of a binary outcome using a \*\*sigmoid function\*\*, which maps the linear combination of features to a probability between 0 and 1. This allows for a clear interpretation of the output as the likelihood of belonging to a particular class. Linear regression, in contrast, uses a linear function to predict a continuous value, which doesn't have a direct interpretation as a probability. The sigmoid function is the key differentiator that makes logistic regression suitable for classification tasks.

Answer 2

The sigmoid function plays a vital role in logistic regression by transforming the output of a linear combination of features into a probability. The linear part of the model, w^T x + b, can produce any real number, which is not suitable for representing probabilities. The sigmoid function, σ(z) = 1 / (1 + e^-z), where z = w^T x + b, addresses this.

Its key mathematical properties are:

\* \*\*Bounded Range (0, 1):\*\* The exponential term in the denominator ensures that the output is always between 0 and 1, which is a fundamental requirement for probabilities.

\* \*\*Differentiability:\*\* It's infinitely differentiable, which is crucial for using gradient-based optimization algorithms like gradient descent to train the model.

\* \*\*Monotonicity:\*\* It's a monotonically increasing function, which means that as the input increases, the output probability also increases.

\* \*\*Smooth S-Shape:\*\* The smooth S-shaped curve allows for a gradual transition between the two extremes (0 and 1), making it suitable for modeling probabilities.

\* \*\*Symmetry around 0.5:\*\* The function is symmetric around 0.5, which means that if the input is 0, the output is 0.5. This is useful for binary classification where 0.5 is often used as a decision threshold.

Answer 3

Response: "In logistic regression, the goal is to model the probability of a binary outcome. The likelihood function measures how well the model's parameters explain the observed data. However, the likelihood can be very small, leading to numerical instability. Therefore, we maximize the log-likelihood instead. Maximizing the log-likelihood is equivalent to maximizing the likelihood, but it's mathematically easier and more stable. It helps us find the model parameters that best fit the training data, meaning that the model assigns high probabilities to the observed outcomes."

Explanation This answer demonstrates a good understanding of the problem, the use of the likelihood function, the need for log-likelihood, and the goal of finding parameters that fit the data well. It also explains the numerical stability aspect. It's a comprehensive answer.