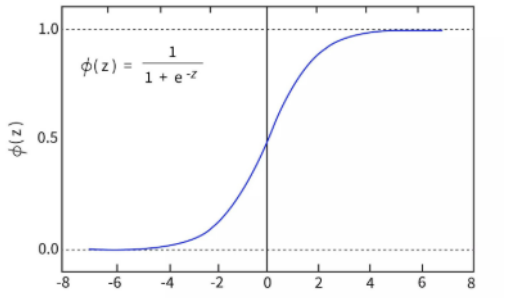
Logistic Regression

Type: Supervised Learning **Task:** Binary Classification

Steps Involved:

- Data Preparation: Organize labeled training data with features and binary outcomes.
- Hypothesis Function: Apply the sigmoid function to map inputs to probabilities.
- Cost Function
- Optimization: Minimize the cost function via gradient descent to update coefficients
- Prediction



Hyperparameters

- C: Regularization strength (inverse of penalty weight).
- solver: Optimization algorithm (e.g., lbfgs, liblinear).
- max_iter: Maximum iterations for convergence.
- penalty: Regularization type (L1/L2).

Assumptions:

- Binary Outcome: Dependent variables must be dichotomous.
- Independence: Observations are independent.
- Linearity: Independent variables relate linearly to log odds.
- No Multicollinearity: Features should not be highly correlated.

Pros	Cons
Simple implementation and interpretation	Struggles with non-linear relationships
Efficient with linearly separable data	Prone to overfitting with high dimensions
Outputs probabilistic predictions	Requires large sample sizes

Example: Imagine predicting if a student will pass a test based on study hours. If they study a lot, the model says "probably yes"; if not, "probably no"-like a magic calculator that guesses outcomes using patterns.

What is logistic regression?: A tool to predict yes/no outcomes using historical data.

How does it differ from linear regression?: Linear regression predicts numbers; logistic predicted probabilities.

What is regularization?: A technique to prevent overfitting by penalizing large coefficients.