

Statistical Classification and Logistic Regression

Introduction to
Parametric and Non-Parametric Models

Purpose

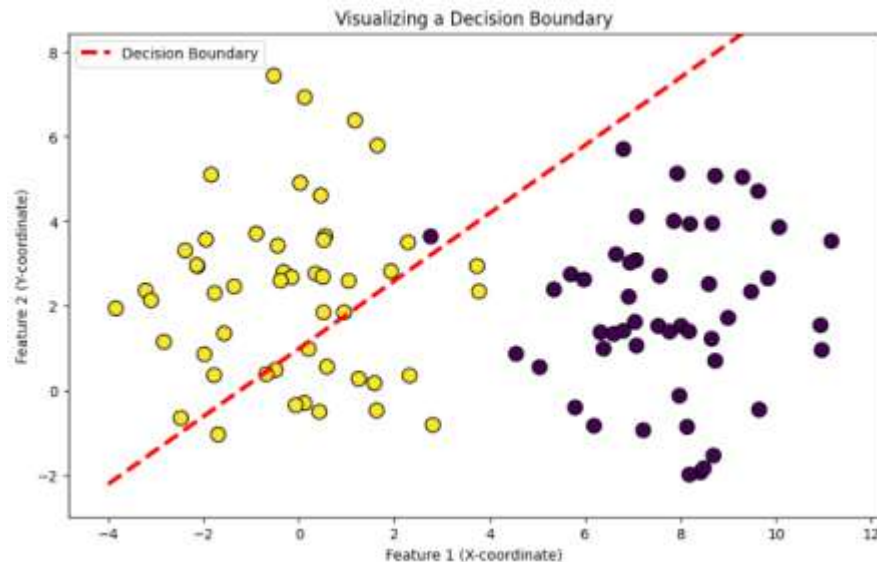
Regression aims to predict continuous output variables based on the input variables.

- The output variables are the dependent variables
- The input variable are the independent variables, also called features

What is Statistical Classification?

Teaching a Computer to Sort and Label Data

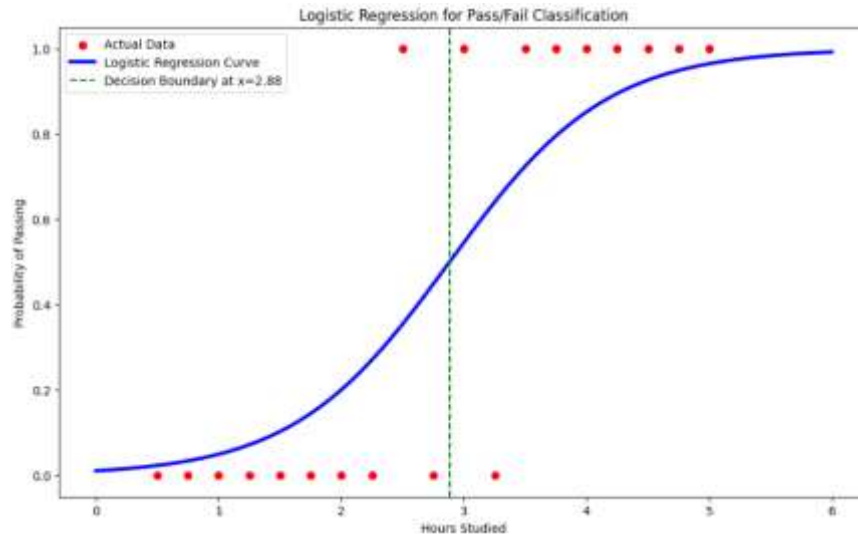
- The Goal: To train a model that can assign a predefined category (or "class") to new, unseen data.
- The Core Question: Based on the evidence (the data), which group does this item belong to?
- Key Ingredients:
 - Features: The measurable pieces of input data used to make a decision. (e.g., for a patient, features could be age, weight, and blood pressure).
 - Labels (or Classes): The predefined categories we are trying to predict. (e.g., Is Diabetic ● vs. Is Not Diabetic ●).



A Tool for Classification: Logistic Regression

Moving from Data to Probabilities

- A Foundational Algorithm: Despite having "regression" in its name, it is the go-to method for binary classification (problems with two classes).
- The Key Idea: Instead of just outputting a "Yes" or "No" answer, Logistic Regression calculates the probability that an input belongs to a class.
 - It gives a value between 0 and 1 (or 0% and 100%).
- How it Works: It uses a special "S-shaped" function called the Sigmoid Curve to turn any input into a probability.
 - A threshold (usually 0.5 or 50%) is then used to make the final classification.



Strategy	Description	How It Adapts to LLMs
Pair Programming	Pairs switch roles between coder and reviewer, or “driver” and “navigator.”	Even if using LLMs, students must <i>discuss</i> and <i>justify</i> choices—builds teamwork and accountability



- 1 Study the algorithm to implement the Statistical Classification Log-Loss (a.k.a., Cross Entropy). Take the #hours studied vs. pass-fail use case as an example. Document the algorithm and functions in a Jupyter Notebook using markdown and code cells.
- 2 Commit and push your solution to your GitHub
- 3 Prepare three talking points about the most important sections of your code

When you are ready:

You

Your peer

- | | |
|--|---|
| <ol style="list-style-type: none"> 4a Clone your peer’s code, review it, and ask questions about the code | <ol style="list-style-type: none"> 4b Will clone your code, review it, and ask questions about it. |
| <ol style="list-style-type: none"> 5 Be ready to present both your notes and reflections <u>as pairs</u> | |