

Task2WriteUp

2024-12-09

Task 2: Feature Selection

Abstract Task 2

The focus of the following analysis is to attempt to achieve high classification accuracy, while using a minimal number of features on very highly dimensional data. To accomplish the following,

Data Description

We were given one dataset that contains binary features that describe the three-dimensional properties of a molecule in a compound or a random probe across different compounds. The data contains 800 observations, which represent 800 compounds, and 100,000 columns (50,000 real variables and 50,000 random probes). The first column named *label* represents whether a compound bound to a target site on thrombin¹.

¹₁