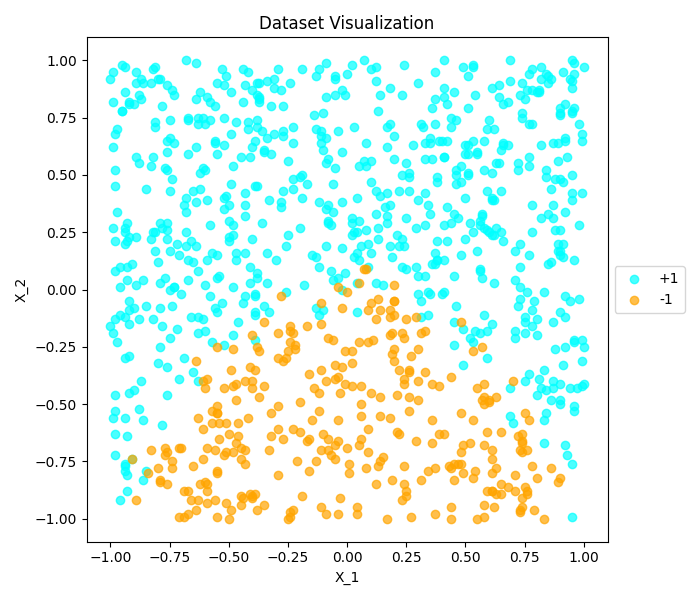
**CS7CS4 Machine Learning**

**Week 2 Assignment Report Priyansh Nayak, 25350660**

**The assigned dataset is: # id:16-16-16**

**Part A**

(i) The provided dataset consists of two numerical features (dubbed X1 and X2) and a binary target label (that was -1 or +1). The features were plotted into a scatter plot, with the X-axis representing the X1 feature and the Y-axis representing the X2 feature. Data points with label as +1 are represented using cyan circles, while those with label as -1 as orange circles for clarity. Here is the graph produced:



*Figure 1. Scatter plot of the given data*

The code segment that produced this is present under the plot\_given\_data() function.

(ii) The Logistic Regression model was trained using both the given features, which then computed the function: which is the sigmoid to predict the class labels (i.e. +1 or -1).

After fitting, the feature coefficients that were produced were and , while the intercept . The accuracy of this model compared to the actual data produced a score of 0.87888 i.e. approximately 87.9%. The interpretation of the coefficients is as follows:

* is negative but small, which implies that increasing feature would slightly decrease the probability of predicting +1.
* is positive and larger, which implies that increasing feature would increase the probability of predicting +1.
* As such, since , feature would have the most influence on prediction.

This analysis aligns with how the data is structured in Figure 1, where the data points appear to be vertically segregated into two distinct regions.