

## CS-677 Assignment: Linear Separability

This report examines the linear separability of weekly stock return data using mean return ( $\mu$ ) and volatility ( $\sigma$ ). The study involves removing overlapping data points, deriving a decision boundary, applying it to new years, and evaluating the impact on portfolio growth.

### 1. Constructing a Reduced Dataset and Finding a Decision Boundary

We started by plotting Year 1's data with  $\mu$  on the x-axis and  $\sigma$  on the y-axis, with green points for 'Green' weeks and red points for 'Red' weeks. We identified points where red and green overlapped significantly and manually removed specific weeks (41, 39, 19, 33, 15, 29, 38) to create a more separable dataset.

After removing these points, we computed a decision boundary by fitting a linear equation between the adjusted centroids of the red and green points. The final equation is:

$$\sigma = 1.40 * \mu + 204.53$$

This decision boundary ensures that most red points remain above the line and most green points stay below.

### 2. Using the Decision Boundary for Classification in Years 2-5

We applied the computed equation ( $\sigma = 1.40 * \mu + 204.53$ ) to classify weeks in Years 2, 3, 4, and 5. If a week's  $\sigma$  was greater than the computed threshold, it was classified as 'Red'; otherwise, it was 'Green'. We then compared the predicted labels with actual labels to measure accuracy.

The classification accuracy for each year was:

- 2021: 78.85%
- 2022: 90.38%
- 2023: 94.23%
- 2024: 90.57%

This suggests that the decision boundary is a reasonable classifier, particularly in later years where accuracy improves.

### 3. Portfolio Growth: Strategy vs. Buy-and-Hold

We simulated an investment of \$100 at the start of each year using two strategies:

- **Buy-and-hold**: Invested throughout the year.
- **Strategy-based investing**: Invested only in predicted 'Green' weeks.

To ensure realistic results, we capped weekly returns to  $\pm 50\%$  to avoid extreme fluctuations. The final portfolio values were:

- **2021**: Strategy = \$77,244.36 | Buy & Hold = \$98.21
- **2022**: Strategy = \$5,766.50 | Buy & Hold = \$0.00

- **2023**: Strategy = \$307,220.41 | Buy & Hold = \$1,922.37
- **2024**: Strategy = \$1,017,140.28 | Buy & Hold = \$535.47

In all years, the strategy significantly outperformed buy-and-hold, especially in volatile years where avoiding 'Red' weeks prevented major losses.

#### 4. Comparing Graphs and Discussion

Key observations from the scatter plots and portfolio growth analysis:

- **Red weeks tend to be in high-volatility zones**, which aligns with our classification method.
- **The decision boundary remains effective across years**, achieving 90%+ accuracy in later years.
- **Avoiding 'Red' weeks consistently improves returns**, proving that selective investment strategies can reduce risk and maximize growth.
- **Buy-and-hold often results in losses or minimal gains**, highlighting the importance of filtering out high-risk weeks.

Overall, the linear classification approach effectively separates volatile periods and offers a strong alternative to passive investing.