# Examine labels:

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**1. Are there any ”obvious” patterns? For example, for higher σ and μ are there more green points?**

There are some patterns. For example, in both of the pictures, the green dots are almost all on the right side (higher μ), and the red dots are on the left side (lower μ). It seems that σ does not determine any pattern here.

**2. are points of the same color located close to each other?**

Yes, most of the points of the same color are close to each other.

**3. do patterns repeat from year 1 to year 2?**

Yes, in both year 1 and year 2, the pattern repeats.

**4. do you expect the nearest-neighbor classifier trained in year 1 to do well in year 2?**

Yes, since the pattern is very close to each other.

# Separate points:

**1. take year 1 and examine the plot of your labels. Construct a reduced dataset by removing some green and red points so that you can draw a line separating the points. Compute the equation of such a line (many solutions are possible)**

Chart, scatter chart

Description automatically generated

**2. take this line and use it to assign labels for year 2**

Chart, scatter chart

Description automatically generated

**3. implement a trading strategy based on your labels for year 2.**

Label red when the daily return is lower than 0.05 and green when the daily return is greater than 0.05.

# Retail Price Data

1. plot 3 histograms for the frequencies for real distribution, equal-weight and Bernford (for each digit)

Chart, bar chart

Description automatically generated

**2. plot 3 histograms for the relative errors for Models 1 and 2 (for each digit)**

Chart, bar chart

Description automatically generated

**3. compute RMSE (root mean squared error) for model 1, 2. Whick model is closer to the real distribution?**

A picture containing text

Description automatically generated

Model 2 is closer to the real distribution.

**4. take 3 countires of your choice: one from Asia, one from Europe and one from the Middle East. For each of these countries do the following:**

**(a) compute F, P and π**

**(b) using RMSE as a ”distance” metric, for which of these chosen three countries is the distribution ”closest” to equal weight P?**

**Text

Description automatically generated with medium confidence**

**Text

Description automatically generated**

Israel is the country that has the closest distance to its RMSE.

**5. discuss your findings**

According to the given data, the Benford distribution is always more accurate in prediction than the equal-weight distribution. If the data sample has been changed, say the sample is more randomized; then the result might vary.