

# Assignment 2 Report

Aarathy Babu, Dewi Amaliah, Priya Dingorkar, Rahul Bharadwaj

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Preliminary Analysis</b>	<b>1</b>
<b>3</b>	<b>Multidimensional Scaling (MDS)</b>	<b>1</b>
3.1	Multidimensional Scaling (MDS) . . . . .	1
3.2	Goodness of Fit . . . . .	3
<b>4</b>	<b>Principal Component Analysis (PCA)</b>	<b>5</b>
<b>5</b>	<b>Cluster Analysis</b>	<b>5</b>
<b>6</b>	<b>Limitations</b>	<b>5</b>
<b>7</b>	<b>Conclusions</b>	<b>5</b>
<b>8</b>	<b>Appendix</b>	<b>5</b>

## 1 Introduction

## 2 Preliminary Analysis

## 3 Multidimensional Scaling (MDS)

```
## # A tibble: 2 x 2
##   SIC      median
##   <fct>   <dbl>
## 1 20      304.
## 2 53      683
```

### 3.1 Multidimensional Scaling (MDS)

MDS is a statistical method to represent multidimensional data into lower-dimensional (2D) data. The `bankruptcy` data has 21 variables which can be considered as data with high dimensions. Thus, MDS is relevant to represent this data in two-dimensional visualisation. This method uses distance to do the job. Hence, we limit the MDS only to incorporate numerical variables so that we can use Euclidean distance or is known as classical MDS. We will also only incorporate numerical variables directly related to `bankruptcy`. Those variables are: `Assest`, `DaysIn`, `Employees`, `CPI`, `Ebit`, `Liab`, `FillingRate`, `GDP`, `PrimeFilling`, and `Sales`. These variables has different unit of measurements, hence we standardise it.

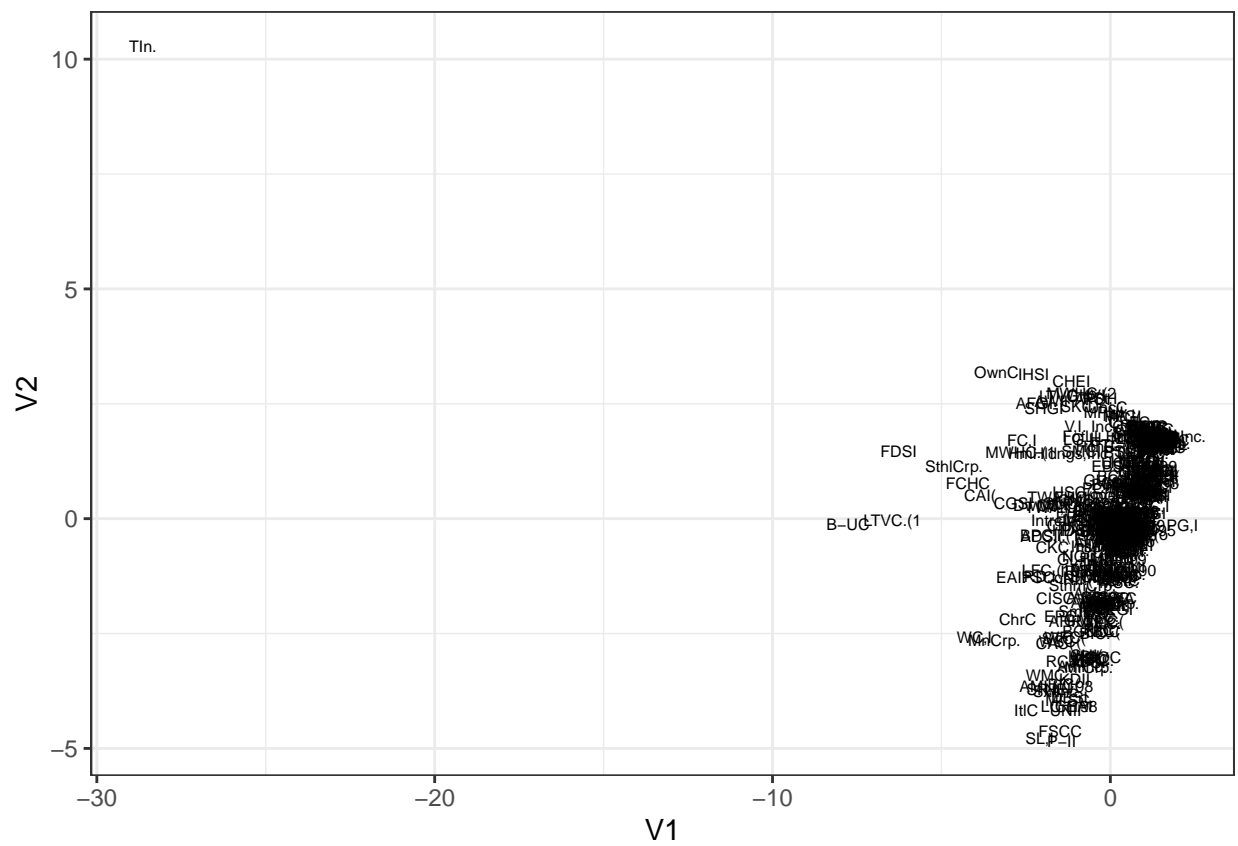


Figure 1: Classical MDS solution for bankruptcy data. The x and y-axis represent the new variables as the result of MDS. Some outliers observed in the data

### 3.1.1 Classical MDS

Figure 1 conveys that Texaco Inc (Tin.), Baldwin-United Corporation (B-UC), Federated Department Stores, Inc. (FDSI), LTV Corp. (1986) (LTCV.(1) are potential outliers. On closer inspection of the data, we find that these firms have the largest assets. Moreover, Texaco Inc. also has high operating income, sales, and liability.

As mentioned previously, the aim of MDS is to visualise the firms in 2D scatter plot. However, this objective will be less clearly achieved in Figure 1 since too many observations overlapped each other. Hence, we decide to exclude Texaco Inc. and re-conduct classical MDS. This gives us a clearer visualisation as follows:

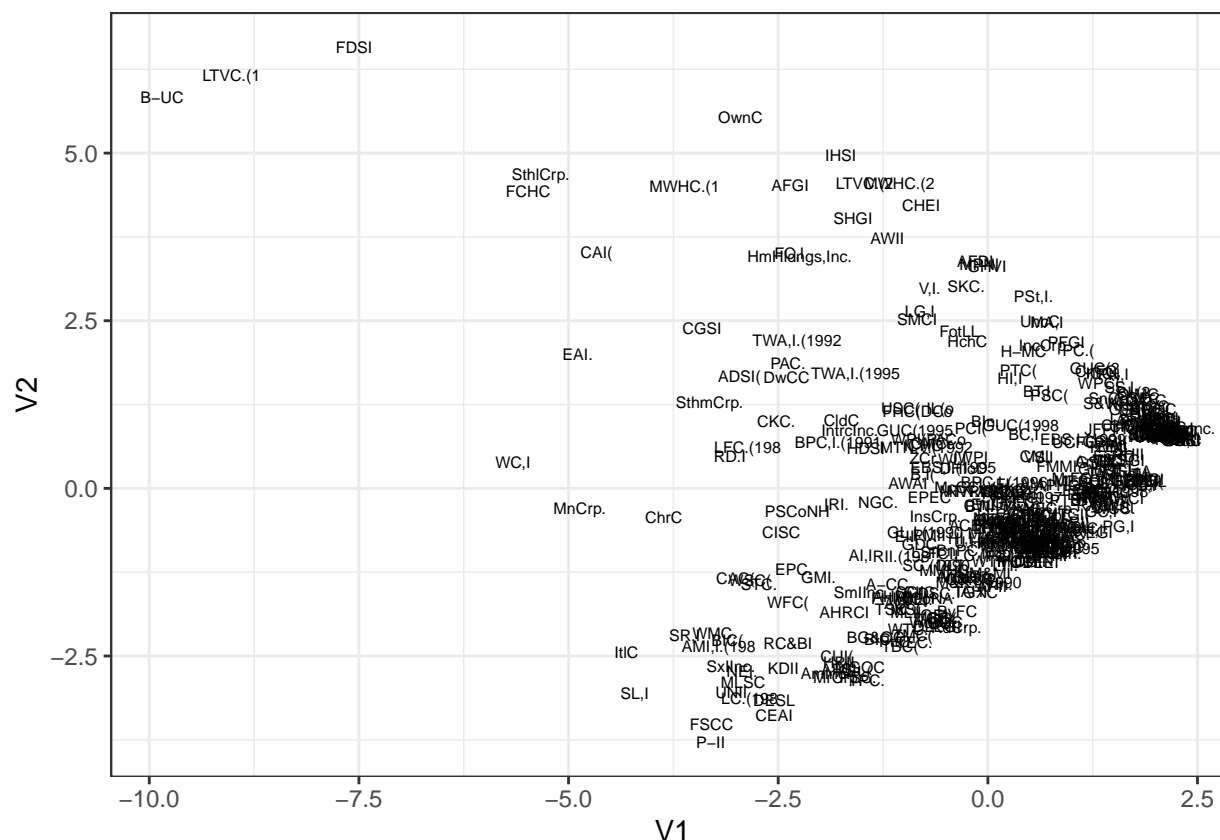


Figure 2: Classical MDS solution for bankruptcy data after excluding Texaco Inc. The x and y-axis represent the new variables as the result of MDS. We get a clearer visualisation compared to the previous MDS result

Figure 2 suggests that the visual representation of the rest firms other than Texaco, Inc. remains the same. B-UC, LTCV.(1, and FDSI are still far apart from other firms. It implies that our MDS is pretty robust. However, since it gives a clearer visualisation, we will use the data without Texaco, Inc. in the rest of MDS analysis. It also implies that most firms that filed for bankruptcy have similar characteristics since they tend to be plotted near or even overlapped with each other. We can also see that some firms are spread out. It means that these firms have different profile.

### 3.2 Goodness of Fit

In this part, we inspect the MDS's Goodness of Fit. If two GoF values are equal, which is the ideal condition if we use Euclidean distance, then we can conclude that the strain is minimised and the solution is optimal. Here is the GoFs of the MDS:

```
## num [1:2] 0.579 0.579
```

We find that the  $GoF_1$  and  $GoF_2$  are equal. Hence, our MDS is optimal. We also find that all the eigenvalues are positive (see Appendix).

### 3.2.1 Comparison with non-Classical MDS

Next, we compare the classical MDS with non-classical MDS (Sammon mapping). The stress function could be used to indicate the accuracy of representation. The lower, the better the accuracy.

```
## Initial stress      : 0.12530
## stress after 2 iters: 0.12098
## [1] 0.1209795
```

We find that the stress is relatively low (0.121), thus non-classical MDS also produce fairly accurate representation of the bankruptcy data. Moreover, the plot (see Appendix) also produce relatively similar result when compared with the classical MDS. Hence, we can conclude that the result is fairly robust with the change of methodology.

### 3.2.2 Visualisation with Categorical Variable

This section will show the MDS solution by also take the categorical variables into account. Too keep the report concise, we display some categorical features in the Appendix and only display interesting finding in this subsection.

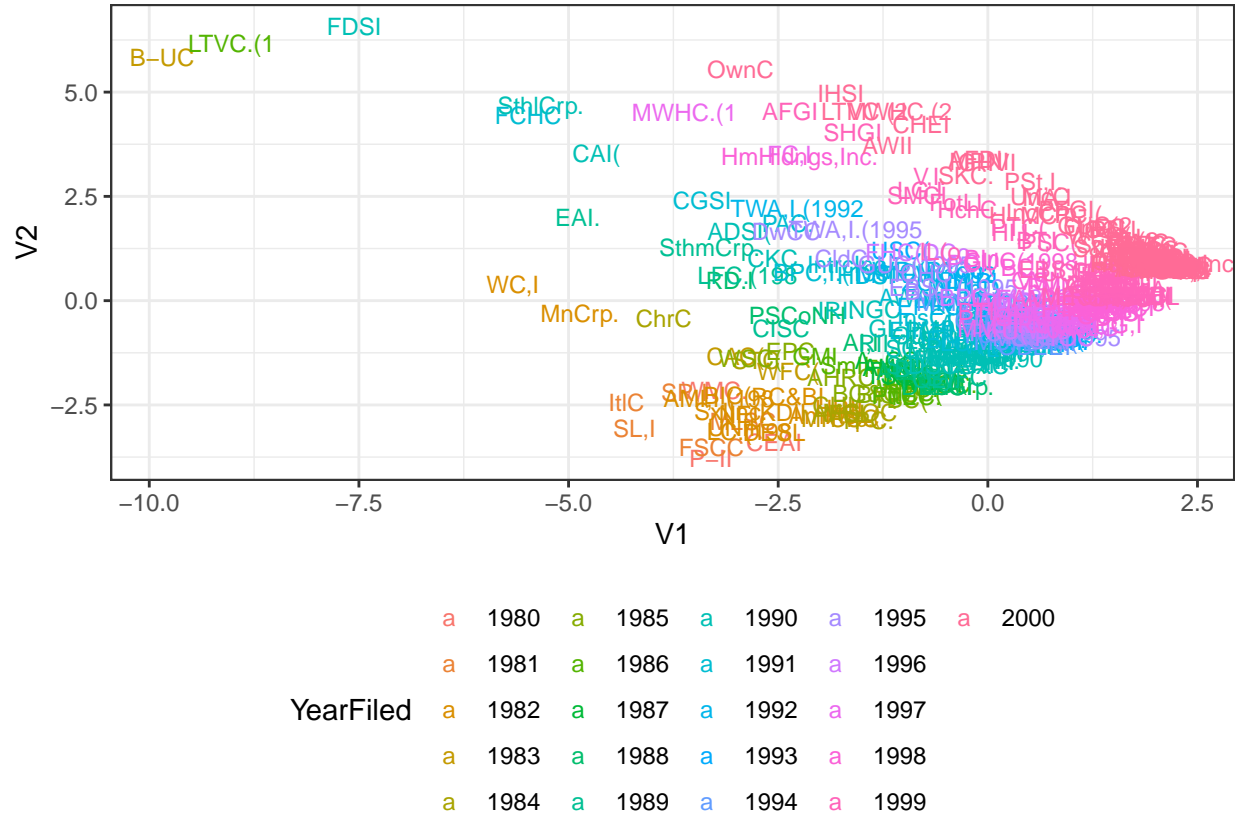


Figure 3: Classical MDS solution plotted by year when the bankruptcy filed.

The classical MDS solution plotted by year as shown in Figure 3 shows that there is pattern regarding the year. Firms who filed for bankruptcy in the same year tend to be similar each other. This could be because

in the same year, CPI, filing rate, and prime interest are pretty similar. This is an interesting finding since we could infer that macroeconomic ,i.e, market condition could profile firms who filed for bankruptcy.

## 4 Principal Component Analysis (PCA)

## 5 Cluster Analysis

## 6 Limitations

- We only use numerical data in the analysis due to the complexity of incorporating non-numeric data. However, we tried to also display that categorical variable when visualising the MDS result.

## 7 Conclusions

## 8 Appendix

### Eigenvalues of classical MDS

```
## [1] -2.065743e-12
```

Since the values has e-12, it is reciprocal to 2 with 12 trailing zeros. Hence, even though it looks negative, it is very close, even indistinguishable from zero. That is why the value of  $GoF_1$  and  $GoF_2$  are equal.

### MDS plot using Sammon mapping

#### Additional plots of MDS based on the city where the bankruptcy filed

Figure 5 shows no specific pattern of bankruptcy regarding the city where it is filed. The firms who similar to each other (as seen in the overlapped text) could filed for bankruptcy in different city. Besides, firms who are potentially outliers (B-UC, LTCV.(1, and FDSI) are not filed their bankruptcy in Delaware.

#### Additional plots of MDS based on industry

Figure 6 shows the classical MDS solution by industry classification. Note that in the original data, there are 55 industry. This number is too big to be plotted, hence we collapse some industry which has the similar sector, for example manufacture, mining, construction, and finance.

Figure 6 suggests that there is no clear specific pattern of the firm bankruptcy regarding the industry. Wholesale and retail firms is bit more spread out. Manufacture industry is also observed to be spread out everywhere and could be because this industry has many observations. Further, B-UC and SthmCrp. are observed to be relatively further apart from the other real estate firms since they have bigger assets.





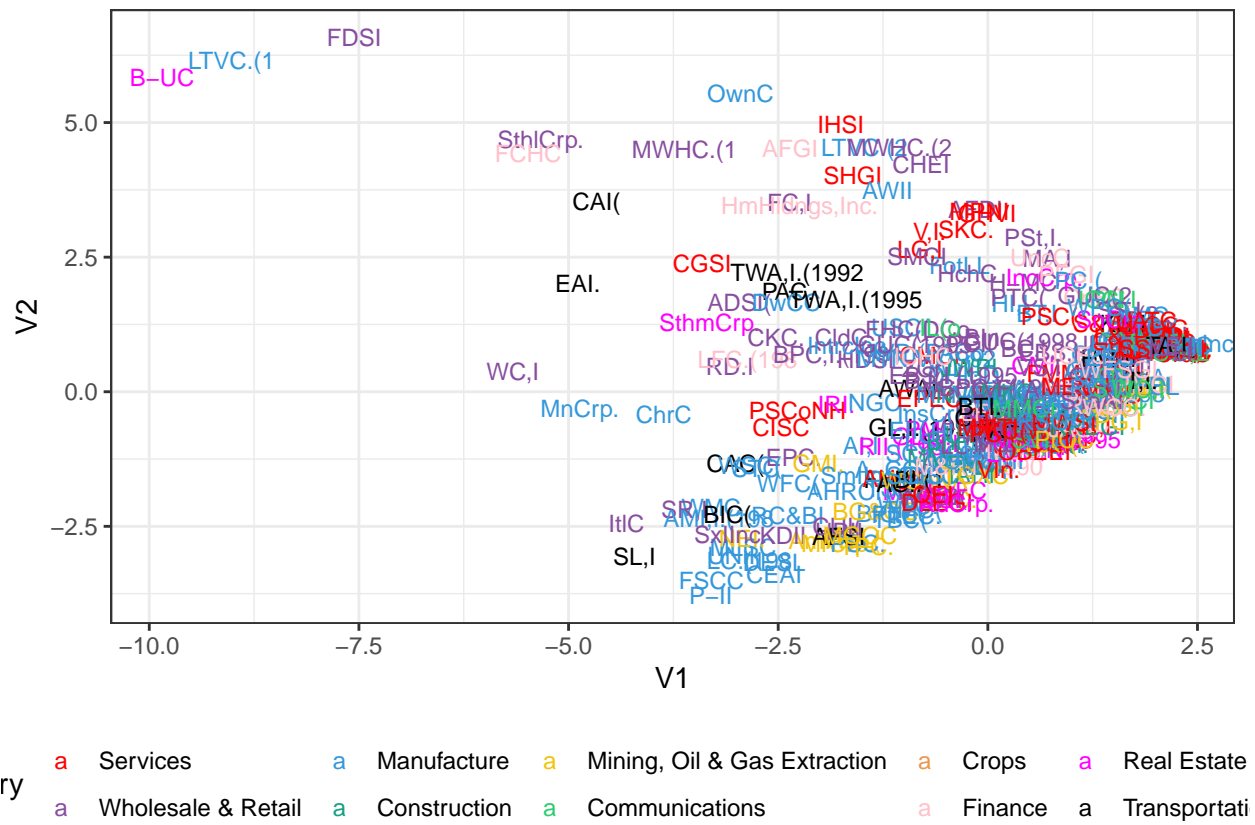


Figure 6: Classical MDS solution plotted by industry.