Introduction on the dataset

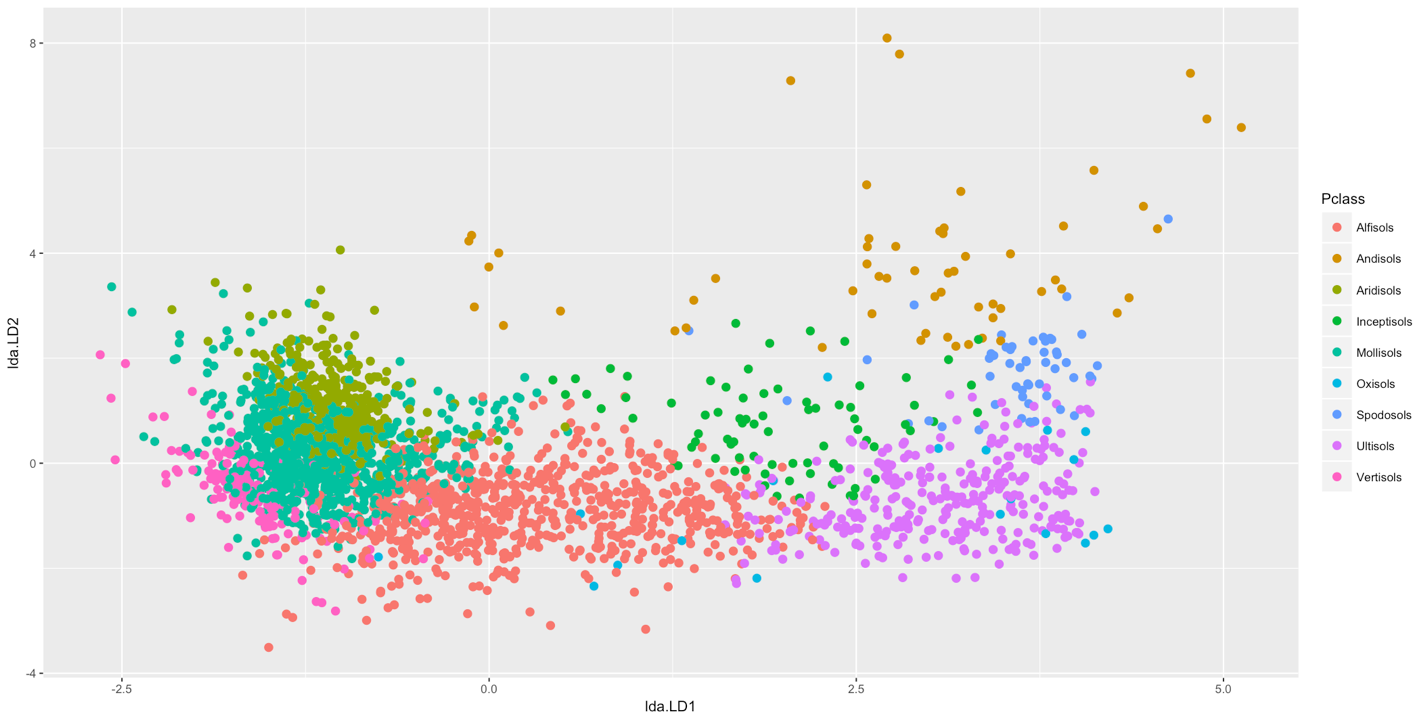
This dataset contains soil sample data over the US downloaded from Natural Resources Conservation Service (NRCS). After removing the incomplete data records, there are around 14,000 records left, each of which includes physical and chemical properties of soil samples (sand, silt, clay, organic carbon, bulk density, CEC soil, CEC clay, base saturation, and pH) and the corresponding soil classification group (soil order).

Linear Discriminant Analysis

Procedure: randomly divided the entire dataset as training data (80%) and testing data (20%) and then fitted LDA model using training data and tested the model using testing data

Results

|  |  |
| --- | --- |
| Soil Class | Accuracy |
| Alfisols | 0.6038462 |
| Andisols | 0.4324324 |
| Aridisols | 0.5673077 |
| Inceptisols | 0.1320755 |
| Mollisols | 0.6547912 |
| Oxisols | 0.5882353 |
| Spodosols | 0.2790698 |
| Ultisols | 0.7758621 |
| Vertisols | 0.6478873 |



Quadratic Discriminant Analysis

Procedure: randomly divided the entire dataset as training data (80%) and testing data (20%) and then fitted QDA model using training data and tested the model using testing data

Results

|  |  |
| --- | --- |
| Soil Class | Accuracy |
| Alfisols | 0.7205128 |
| Andisols | 0.2972973 |
| Aridisols | 0.5432692 |
| Inceptisols | 0.2311321 |
| Mollisols | 0.5159705 |
| Oxisols | 0.8823529 |
| Spodosols | 0.4069767 |
| Ultisols | 0.7844828 |
| Vertisols | 0.7183099 |