Data available with the LDR package

1 Classification data

Iris

- Source:
  `http://archive.ics.uci.edu/ml/datasets/Iris`
- File: `iris.txt`
- Description: This is a classical data set from a paper by Fisher. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant.
- Response: $Y = \{1, 2, 3\}$ the types of iris plant, relocated in the first column of the data file (it is in the last column in the original data file).
- Predictors: sepal and petal length and width, relocated in columns 2 to 5 of the file.

Handwritten digits

- Source:
  `ftp://ftp.ics.uci.edu/pub/machine-learning-databases/pendigits`
- File: `digits_full.txt` for all the digits and `digits.txt` for digits 0, 6 and 9 only.
- Description: 44 subjects were asked to write 250 random digits. Each digit yields a 16-dimensional feature vector, consisting of 8 pairs of randomly sampled two-dimensional locations on the digit. The 44 subjects were divided into two groups of size 30 and 14, in which the first formed the training set with sample size 7,494 and the second formed the test set with sample size 3,498.
- Response: $Y = \{0, 1, \ldots, 9\}$, relocated in the first column of the file (it is in the last column of the original data file).
- Predictors: columns 2 to 17 of the data file.

Birds, cars and planes

- File: `marcewithout.txt`
- Description: This data set concerns the identification of the sounds made by birds, planes and cars. A two hour recording was made in the city of Ermont, France, and then 5 second snippets of interesting sounds were manually selected. This resulted in 58 recordings identified as birds, 44 as cars and 67 as planes. Each recording was further processed, and was ultimately represented by 13 SDMFCCs (Scale Dependent Mel- Frequency Cepstrum Coefficients).
- Response: $Y = \{1, 2, 3\}$ refering to birds, planes or cars. It is located in the first column of the data file.
- Predictors: 13 scale dependent Mel-frequency cepstrum coefficients, located in columns 2 to 14 of the data file.

Bank notes

- File: `banknotes.txt`
- Description: Six variables measured on 100 genuine and 100 counterfeit old Swiss 1000-franc bank notes. The data stem from Flury and Riedwyl (1988). Each feature gives a different aspect of the size of a note.
- Response: $Y = \{0, 1\}$ indicating genuine or counterfeit bank note, respectively, located in the first column of the data file.
- Predictors: located from column 2 to 7 of the data file.

Wine

- Source:
  `http://archive.ics.uci.edu/ml/datasets/Wine`
- File: `wine.txt`
- Description: These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.
- Response: $Y = \{1, 2, 3\}$ the types of wine, located in the first column of the data file.
- Predictors: 13 chemical attributes of wines located in columns 2 to 14 of the data file.

2 Regression data

Naphtalene

- File: `nap.txt`
- Description: This data consist of 80 observations from an experiment to study the effects of three process variables in the vapor phase oxidation of naphthalene. The response is the percentage conversion of naphthalene to naphthoquinone, and the three predictors are air to naphthalene ratio, contact time and bed temperature. Although there are only three predictors in this regression, dimension reduction may still be useful for visualization.
- Response: percentage conversion of naphthalene to naphthoquinone.
- Predictors: air to naphthalene ratio, contact time and bed temperature.

Fearn's wheat protein

- File: `fearns.txt`
- Description: This is Fearn's calibration data that contains the protein content of a sample of ground wheat, and minus -log(reflectance) of NIR radiation at six wavelengths as predictors.
- Response: Protein content, located in the first column of the data file.
- Predictors: -log(reflectance) of NIR radiation at six different wavelengths, located in columns 2 to 7 of teh data file.

Horse Mussel

- File: `mussels.txt`
- Description: Horse mussels data collected at 5 sites in Marlborough Sounds, in New Zeland. The response variable is muscle mass in grams. Predictors concern characteristics of the mussel shell.
- Response: muscle mass in grams, located in the first column of the data file.
- Predictors: four characteristics of the mussel shell, located from column 2 to 5. Indicator predictors for the site of collection are not included.