**The Landsat data**

* **Description**: This dataset contains pixel information for a 82\*100 pixel part of an image of the earth surface taken from a satellite. For each pixel, 4 frequency values (each between 0-255) are given. This is because pictures are taken in 4 different spectral bands. Each pixel represents an area on the earth's surface of 80\*80 metres. Every line in the dataset gives the 4 values for 9 pixels: a 3\*3 neighbourhood frame around a central pixel. The pixels are numbered from left to right and from top to bottom, so pixel 1 is the left top one, pixel 3 is the right top one, and pixel 5 is the central one. More information about this dataset is given in the [documentation file](http://www.inf.ed.ac.uk/teaching/courses/dme/html/satdoc.txt).
* **Source**: This dataset was obtained from [the Statlog data collection](http://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29) in the [UCI Machine Learning Repository](http://archive.ics.uci.edu/ml/).
* **Size**:
  + 4435 training cases and 2000 test cases
  + 36 attributes and 1 class attribute
  + 514K training data and 231K test data
* **Task**: The aim of this dataset is to classify the central pixel (pixel 5) into one of 6 classes, based on its four frequency values and those of its 8 immediate neighbours. The classes are red soil (class 1), cotton crop (class 2), grey soil (class 3), damp grey soil (class 4), soil with vegetation stubble (class 5), and very damp grey soil (class 7). Originally there was also a mixture class, where all types are present (class 6), but there are no cases of it in the dataset.