Datasets descriptions

1 Satellite Image Dataset

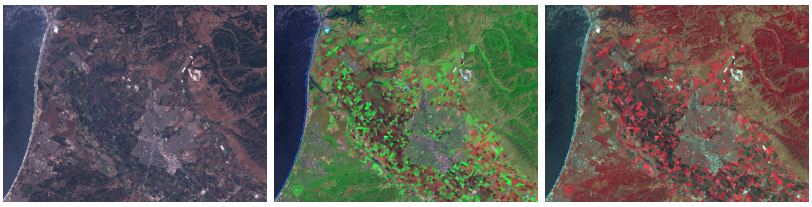

Figure 1: Example of a landsat MSS image. True image on the left and two of the corresponding spectral bands.

One frame of Landsat multi-spectral scanner (MSS) imagery consists of four digital images of the same scene in different spectral bands. Two of these are in the visible region (corresponding approximately to green and red regions of the visible spectrum) and two are in the (near) infra-red. Each pixel is a 8-bit binary word, with 0 corresponding to black and 255 to white. The spatial resolution of a pixel is about 80m x 80m. Each image contains 2340 x 3380 such pixels.

The database is a (tiny) sub-area of a scene, consisting of 82 x 100 pixels. Each line of data corresponds to a 3x3 square neighbourhood of pixels completely contained within the 82x100 sub-area. Each line contains the pixel values in the four spectral bands (converted to ASCII) of each of the 9 pixels in the 3x3 neighbourhood and a number indicating the classification label of the central pixel. The number is a code for the following classes:

1 red soil

2 cotton crop

3 grey soil

4 damp grey soil

5 soil with vegetation stubble

6 mixture class (all types present)

7 very damp grey soil

Here is how examples in this dataset look like:

```
[...]
67 84 85 69 70 88 93 73 74 88 89 73 72 81 86 68 80 98 106 83 88 [...] 106 111 87 68 3
82 92 93 83 78 92 101 80 78 92 97 76 80 98 102 83 80 94 102 79 76 [...] 85 90 68 84 7
78 92 101 80 78 92 97 76 67 71 78 62 80 94 102 79 76 85 90 68 68 77 [...] 90 68 76 87 3
74 79 89 73 78 92 97 87 78 97 101 83 76 85 98 79 76 85 98 79 80 94 [...] 102 83 76 83 1
78 92 97 87 78 97 101 83 82 102 105 87 76 85 98 79 80 94 102 83 [...] 88 106 106 87 7
78 97 101 83 82 102 105 87 85 106 114 90 80 94 102 83 88 106 106 [...] 87 88 106 111 5
82 102 105 87 85 106 114 90 93 120 119 97 88 106 106 87 88 106 [...] 111 91 88 115 2
85 106 114 90 93 120 119 97 93 115 124 97 88 106 111 91 88 115 [...] 120 94 84 111 2
[...]
```

In each line of data the four spectral values for the top-left pixel are given first (in red) followed by the four spectral values for the top-middle pixel (in blue) and then those for the top-right pixel (in green), and so on with the pixels read out in sequence left-to-right and top-to-bottom. Thus, the four spectral values for the central pixel are given by attributes 17,18,19 and 20.

Number of Instances/examples: 6435

- `satimage_train_*.txt` contains 4435 examples for training and validation.
- `satimage_test.txt` contains 2000 examples for testing