

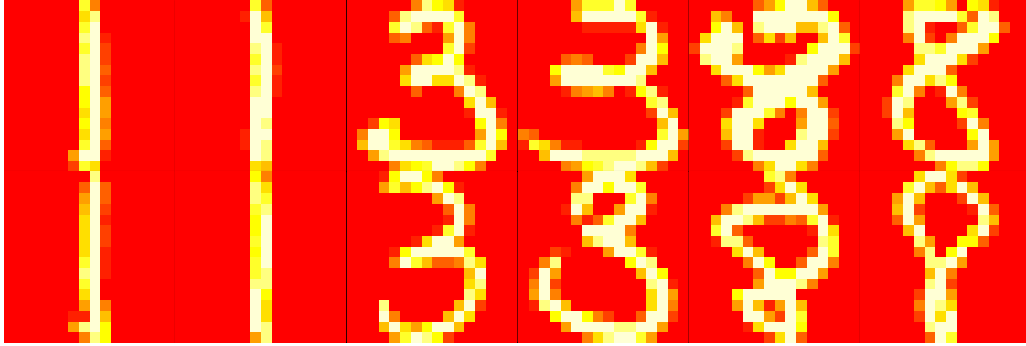
## G14SML dataset: zip code digit data

These data are a subset of data from those used in Elements of Statistical learning, and available on the book's web page, under the “Data” tab. See the info file there for further details.

The data are 16-pixel-by-16-pixel digital images of handwritten digits.

In this cut-down version of the dataset I include only images of the digits 1, 3 and 8.

Some examples of the images:



Load the data with

```
> load("zipCode138.RData")
```

Variables:

- **train.X** is the  $n$ -by- $p$  data matrix. The observations are vectors of length  $p = 256$  ( $= 16 \times 16$ ), the elements of which represent pixel intensities
- **train.X.array** are the same data but rearranged into a 16-by-16- $n$  array<sup>1</sup>. This format is more convenient for feeding into the function `image` to visualise an observation. For example `image(train.X.array[, , 33])` shows the 33rd observation.
- **y** contains the observation labellings, i.e., one of 1, 3 or 8.

---

<sup>1</sup>The specific command I used was `train.X.array <- array(as.numeric(t(train.X)),c(16,16,dim(train.X)[1]))`