

| | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.9 | 2.0 | 1.9 | 2.0 | 2.0 | 2.0 | 1.9 | 2.0 | 1.9 | 1.9 | 2.0 | 2.0 | 2.0 | 2.0 |
| | 1.6 | 2.0 | 1.8 | 2.0 | 2.0 | 2.0 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 2.0 | 1.7 |
| | | 1.5 | 2.0 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 2.0 |
| | | | 1.4 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 1.8 |
| | | | | 1.5 | 1.8 | 1.7 | 1.9 | 1.9 | 1.8 | 1.9 | 1.9 | 1.8 | 2.0 |
| | | | | | 0.9 | 1.7 | 1.9 | 1.9 | 1.8 | 1.9 | 1.9 | 1.9 | 2.0 |
| | | | | | | 0.3 | 1.9 | 1.9 | 1.5 | 1.9 | 1.9 | 1.9 | 2.0 |
| | | | | | | | 1.8 | 1.8 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 |
| | | | | | | | | 1.7 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 |
| | | | | | | | | | 1.6 | 1.9 | 1.9 | 1.9 | 2.0 |
| | | | | | | | | | | 1.8 | 1.9 | 1.9 | 1.9 |
| | | | | | | | | | | | 1.8 | 2.0 | 1.9 |
| | | | | | | | | | | | | 1.5 | 2.0 |
| | | | | | | | | | | | | | 1.5 |

TABLE 12.1: Each column of the table represents an activity for the activity dataset <https://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer>, as does each row. In each of the upper diagonal cells, I have placed the average chi-squared distance between histograms of examples from that pair of classes (I dropped the lower diagonal for clarity). Notice that in general the diagonal terms (average within class distance) are rather smaller than the off diagonal terms. This quite strongly suggests we can use these histograms to classify examples successfully.

PROGRAMMING EXERCISES

- 12.1.** Obtain the activities of daily life dataset from the UC Irvine machine learning website (<https://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer>) data provided by Barbara Bruno, Fulvio Mastrogiovanni and Antonio Sgorbissa).
- (a) Build a classifier that classifies sequences into one of the 14 activities provided. To make features, you should vector quantize, then use a histogram of cluster centers (as described in the subsection; this gives a pretty explicit set of steps to follow). You will find it helpful to use hierarchical k-means to vector quantize. You may use whatever multi-class classifier you wish, though I'd start with R's decision forest, because it's easy to use and effective. You should report (a) the total error rate and (b) the class confusion matrix of your classifier.
 - (b) Now see if you can improve your classifier by (a) modifying the number of cluster centers in your hierarchical k-means and (b) modifying the size of the fixed length samples that you use.