CSC411: Project #2

Due on Sunday, February 18, 2018

Hao Zhang & Tyler Gamvrelis

February 4, 2018

Foreword

In this project,...

System Details for Reproducibility:

- Python 2.7.8
- Libraries:
  - numpy
  - matplotlib
  - ...

Part 1

Dataset description
The dataset is made of thousands of 28 by 28 pixel images of the handwritten digits: 0 to 9. The images are split into training set and test set images labelled 'train0' to 'train9' and 'test0' to 'test9'. The number of images with each label is as follows:

Label	Number of Images	Label	Number of Images
train0	5923	test0	980
train1	6742	test0	1135
train2	5958	test0	1032
train3	6131	test0	1010
train4	5842	test0	982
train5	5421	test0	892
train6	5918	test0	958
train7	6265	test0	1028
train8	5851	test0	974
train9	5949	test0	1009

Ten images of each number were taken from the training sets and displayed in Figure 1. The correct labels of most of the pictures can be descerned at a glance by humans However, since the digits are handwritten, some of them may not be completely obvious. For example, Figure 1cv is categorized as a 9 but looks like an 8.

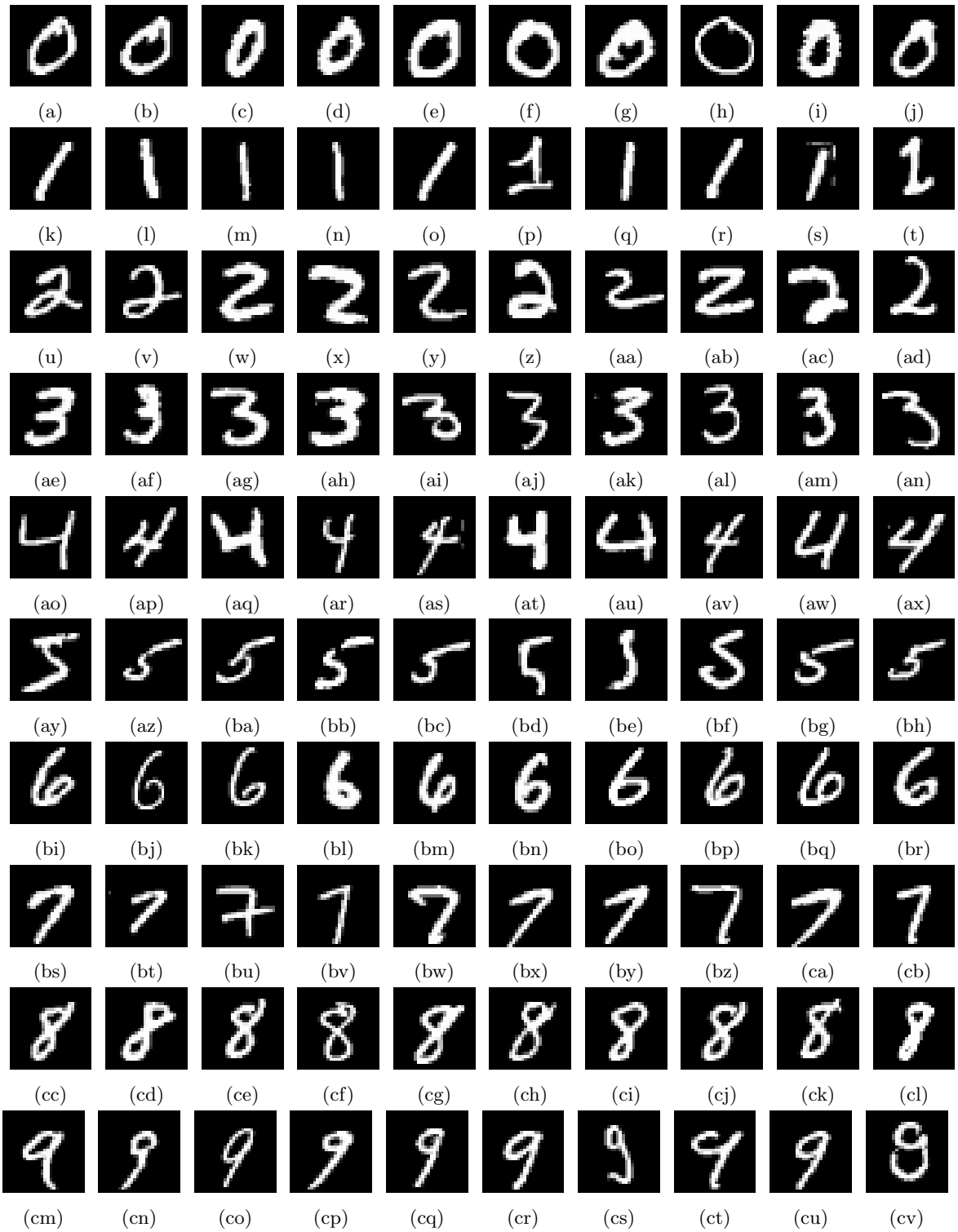
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)
(k) (l) (m) (n) (o) (p) (q) (r) (s) (t)
(u) (v) (w) (x) (y) (z) (aa) (ab) (ac) (ad)
(ae) (af) (ag) (ah) (ai) (aj) (ak) (al) (am) (an)
(ao) (ap) (aq) (ar) (as) (at) (au) (av) (aw) (ax)
(ay) (az) (ba) (bb) (bc) (bd) (be) (bf) (bg) (bh)
(bi) (bj) (bk) (bl) (bm) (bn) (bo) (bp) (bq) (br)
(bs) (bt) (bu) (bv) (bw) (bx) (by) (bz) (ca) (cb)
(cc) (cd) (ce) (cf) (cg) (ch) (ci) (cj) (ck) (cl)
(cm) (cn) (co) (cp) (cq) (cr) (cs) (ct) (cu) (cv)

Figure 1