Deep Self-Taught Learning for Handwritten Character Recognition

Y. Bengio et al @ U. Montreal

- discriminate 62 character classes (upper, lower, digits), 800k to 80M examples
- Deep learners beat state-of-the-art on NIST and reach human-level performance
- Deep learners benefit more from perturbed (out-of-distribution) data
- Deep learners benefit more from multi-task setting

Prediction of Deep Network:

Z — 24 %
z — 8 %
8 — 5 %

Prediction of Shallow Network:

m — 15 %
3 — 8 %
r — 6 %

Relative improvement from perturbed data

□ NIST
■ NIST digits

SDA1   SDA2   MLP1   MLP2

Relative improvement from multi-task setting

SDA-digits   SDA-lower   SDA-upper   MLP-digits   MLP-lower   MLP-upper

NIST 62 character classes (humans, MLPs, SDAs)

% error rate +/- 95% confidence

□ NIST
□ NISTP

Humans   SDA0   SDA1   SDA2   MLP0   MLP1   MLP2

Model

NIST digits, machines vs humans

% error rate +/- 95% confidence

Humans   SDA1   MLP0   [20]   [21]   [22]   [23]

Model