CS498 AML

Ling-Hsi Liu

lhliu2@illinois.edu

Problem 2

Part A
MNIST using naïve Bayes

Accuracy	Gaussian	Bernoulli
Untouched images	53.52 %	60.45%
Stretched Bounding Box	82.84%	83.08%

1. Which is distribution (Gaussian or Bernoulli) is better for untouched pixels?
   The Bernoulli is better.
2. Which is better for stretched bounding box image?
   They are almost the same, but Bernoulli is higher.

Part B
MNIST using a decision forest

Untouched raw pixels:

	Depth = 4	Depth = 8	Depth = 16
#tree =10	85.7%	93.12%	95.67%
#tree =20	86.42%	93.84%	96.17%
#tree = 30	86.77%	94.13%	96.52%

In Untouched raw pixels, the tree is higher. Then, the accuracy is higher.
Moreover, the depth is higher. Then, the accuracy is higher.
Compare two factor, when the depth increases, the accuracy increases more than tree change.

Stretched Bounding Box:

	Depth = 4	Depth = 8	Depth = 16
#tree =10	81.57%	92.18%	95.58%
#tree =20	83.45%	92.85%	96.37%
#tree = 30	84.19%	93.03%	96.62%

In the stretched Bounding Box, the tree and depth is higher, then, the accuracy is higher. However, compared untouched raw pixels and stretched bounding box, the accuracy of untouched raw pixel is higher than the accuracy of stretched bounding box.

Reference:
1. https://gist.github.com/primaryobjects/b0c8333834debbc15be4
2. https://piazza.com/class/jchzguhsowz6n9?cid=247

3. https://piazza.com/class/jchzguhsowz6n9?cid=200
4. https://www.rdocumentation.org/packages/h2o/versions/3.16.0.2/topics/h2o.randomForest