

Handwritten answer sheet auto-grader

Kishore

Hand written digits

- ❖ The handwritten digits are not always of the same size, width, orientation, and they differ from the writing of person to person.
 - ❖ Many people write a single digit with a variety of different handwritings.
 - ❖ The uniqueness and variety in the handwriting of different individuals also influence the formation and appearance of the digits.
-
- **Math exam answer sheets from Turtle Rock Elementary School, Irvine, CA.**

Subtraction

$7-1= \underline{\quad}$

$7-5= \underline{\quad}$

$7-7= \underline{\quad}$

$7-2= \underline{\quad}$

$7-6= \underline{\quad}$

$7-0= \underline{\quad}$

$7-3= \underline{\quad}$

$7-7= \underline{\quad}$

$7-5= \underline{\quad}$

$7-4= \underline{\quad}$

Criterion-Referenced Competency Tests
Answer Key – Grade 8
Form 18
Spring 2000

Reading CRCT

Question	Answer	Question	Answer	Question	Answer
1	C	21	B	31	C
2	B	22	A	32	A
3	A	23	B	33	D
9	D	24	C	34	C
10	B	25	A	35	D
11	B	26	A	36	B
12	A	27	C	47	D
18	B	28	D	48	C
19	C	29	B	49	D
20	B	30	B	50	B

Data description

- Relevant digit records: MNIST Train and test data
 - MNIST: (Modified) National Institute of Standards and Technology database
 - It contains handwritten digits in the form of images.
-
- ✓ Train data set: 60,000 x 787
 - ✓ Test data set: 10,000 x 787

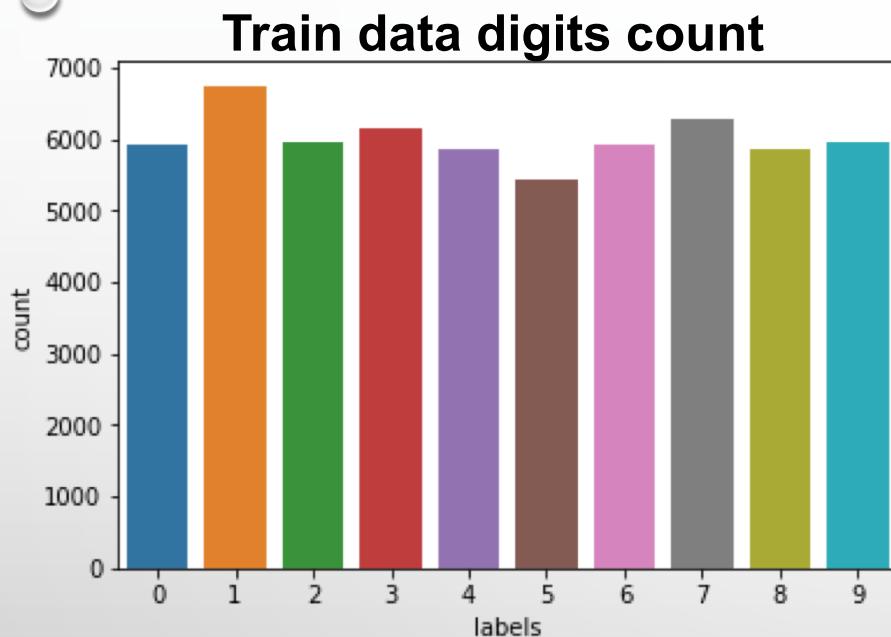
□ Data header information

- Unnamed
- Index
- Labels (The actual true digit that the handwriting is supposed to classify)
- 0, 1, 2, 3, ... 783 ($784 = 28 \times 28$)

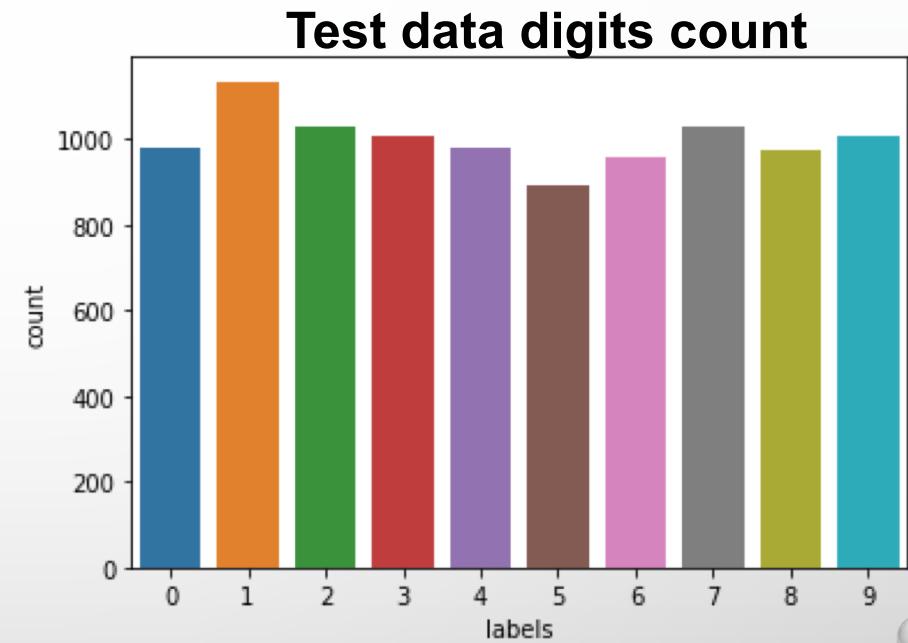
(The pixel values intensities of the handwritten digit, varying from 0 to 255)

HISTOGRAM PLOTS

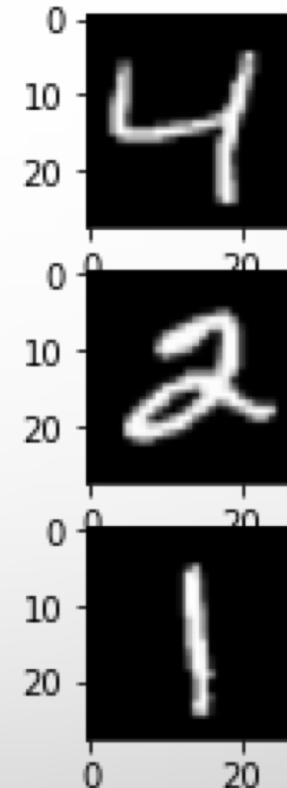
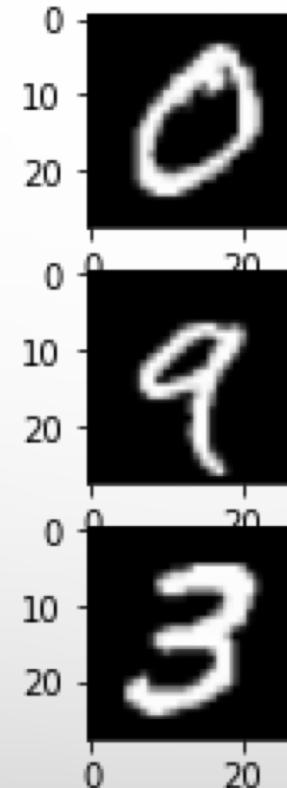
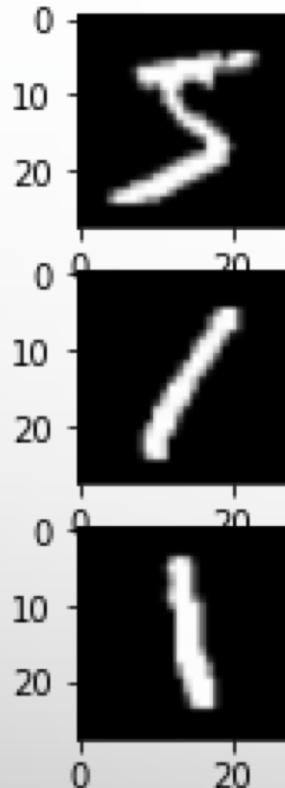
Train data: **85.72%**



Test data: **14.28%**



Data Images 3x3 plot



CLASSIFICATION ALGORITHMS

- 1) Gauss Bayes**
- 2) K Nearest Neighbors (KNN)**

RESULTS:

Gauss Bayes accuracy: 0.9108

K Nearest Neighbors: 0.9633

Conclusions

Parameters	Gauss Bayes	KNN
Accuracy	91.08 %	96.33 %
Time approx.	<3 minutes	40-50 minutes

- KNN accuracy is higher than Gauss Bayes.
- Best choice of K depends on data
- The larger value of K reduce the effect of noise on classification
- If we reduce the noise using PCA or IMD technique, then it will improve the Gauss Bayes accuracy.

*Thank
You*