

ML Homework2

Due Date : 2019/10/14 (MON.) 23:55

TAs' email address: jhhlab.tw@gmail.com

Description :

1. Naive Bayes classifier

Create a Naive Bayes classifier for each handwritten digit that support **discrete** and **continuous** features.

- Input:
 1. Training image data from [MNIST](#)
 - You Must download the MNIST from this website and parse the data by yourself. (Please do not use the build in dataset or you'll not get 100.)
 - Please read the description in the link to understand the format.
 - Basically, each image is represented by $28 \times 28 \times 8$ bits (Whole binary file is in **big endian** format; you need to deal with it), you can use a `char` array to store an image.
 - There are some headers you need to deal with as well, please read the link for more details.
 2. Training lable data from MNIST.
 3. Testing image from MNIST
 4. Testing label from MNIST
 5. Toggle option
 - 0: discrete mode
 - 1: continuous mode

TRAINING SET IMAGE FILE (train-images-idx3-ubyte)

offset	type	value	description
0000	32 bit integer	0x00000803(2051)	magic number
0004	32 bit integer	60000	number of images
0008	32 bit integer	28	number of rows
0012	32 bit integer	28	number of columns
0016	unsigned byte	??	pixel
0017	unsigned byte	??	pixel
...
xxxx	unsigned byte	??	pixel

TRAINING SET LABEL FILE (train-labels-idx1-ubyte)

offset	type	value	description
0000	32 bit integer	0x00000801(2049)	magic number
0004	32 bit integer	60000	number of items
0008	unsigned byte	??	label
0009	unsigned byte	??	label
...
xxxx	unsigned byte	??	label

The labels values are from 0 to 9.

- Output:
 - Print out the the posterior (in log scale to avoid underflow) of the ten categories (0-9) for each image in INPUT 3. Don't forget to marginalize them so sum it up will equal to 1.
 - For each test image, print out your prediction which is the category having the highest posterior, and tally the prediction by comparing with INPUT 4.
 - Print out the imagination of numbers in your Bayes classifier
 - For each digit, print a 28×28 binary image which 0 represents a white pixel, and 1 represents a black pixel.
 - The pixel is 0 when Bayes classifier expect the pixel in this position should less than 128 in original image, otherwise is 1.
 - Calculate and report the error rate in the end.
- Function:
 1. In Discrete mode:

- Tally the frequency of the values of each pixel into 32 bins. For example, The gray level 0 to 7 should be classified to bin 0, gray level 8 to 15 should be bin 1 ... etc. Then perform Naive Bayes classifier. **Note** that to avoid empty bin, you can use a pseudocount (such as the minimum value in other bins) for instead.

- Use MLE to fit a Gaussian distribution for the value of each pixel. Perform Naive Bayes classifier.

[illegible]

```
40 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
41 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 0 0 0 0 0
42 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0
43 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0
44 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0
45 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0
46 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0
47 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
48 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0
49 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0
50 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0
51 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
52 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
53 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
54 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
55 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
56 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
57 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
58 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
59 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

60

61 ... all other imagination of numbers goes here ...

62

63 9:

```
64 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
65 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
66 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
67 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
68 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
69 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
70 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
71 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
72 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
73 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0
74 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0
75 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0
76 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0
77 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
78 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
79 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
80 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0
81 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0
82 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
83 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
84 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
85 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
86 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
87 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
88 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```

89 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
90 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
91 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
92
93 Error rate: 0.1535

```

2. Online learning

Use online learning to learn the beta distribution of the parameter p (chance to see 1) of the coin tossing trails in batch.

- Input:

1. A file contains many lines of binary outcomes:

```

1 0101010111011011010101
2 0110101
3 010110101101

```

2. parameter a for the initial beta prior

3. parameter b for the initial beta prior

- Output: Print out the Binomial likelihood (based on MLE, of course), Beta prior and posterior probability (parameters only) for each line.
- Function: Use Beta-Binomial conjugation to perform online learning.
- Sample input & output (*for reference only*)
 - Input: A file (here shows the content of the file)

```

1 $ cat testfile.txt
2 0101010101001011010101
3 0110101
4 010110101101
5 0101101011101011010
6 111101100011110
7 101110111000110
8 1010010111
9 11101110110
10 01000111101
11 110100111
12 01101010111

```

- Output

- Case 1: $a = 0, b = 0$

```

1 case 1: 0101010101001011010101
2 Likelihood: 0.16818809509277344
3 Beta prior:      a = 0 b = 0
4 Beta posterior: a = 11 b = 11

```

```
5
6 case 2: 0110101
7 Likelihood: 0.29375515303997485
8 Beta prior:      a = 11  b = 11
9 Beta posterior: a = 15  b = 14
10
11 case 3: 010110101101
12 Likelihood: 0.2286054241794335
13 Beta prior:      a = 15  b = 14
14 Beta posterior: a = 22  b = 19
15
16 case 4: 0101101011101011010
17 Likelihood: 0.18286870706509092
18 Beta prior:      a = 22  b = 19
19 Beta posterior: a = 33  b = 27
20
21 case 5: 1111011000111110
22 Likelihood: 0.2143070548857833
23 Beta prior:      a = 33  b = 27
24 Beta posterior: a = 43  b = 32
25
26 case 6: 1011101110001110
27 Likelihood: 0.20659760529408
28 Beta prior:      a = 43  b = 32
29 Beta posterior: a = 52  b = 38
30
31 case 7: 1010010111
32 Likelihood: 0.250822656000000003
33 Beta prior:      a = 52  b = 38
34 Beta posterior: a = 58  b = 42
35
36 case 8: 11101110110
37 Likelihood: 0.2619678932864457
38 Beta prior:      a = 58  b = 42
39 Beta posterior: a = 66  b = 45
40
41 case 9: 01000111101
42 Likelihood: 0.23609128871506807
43 Beta prior:      a = 66  b = 45
44 Beta posterior: a = 72  b = 50
45
46 case 10: 110100111
47 Likelihood: 0.27312909617436365
48 Beta prior:      a = 72  b = 50
49 Beta posterior: a = 78  b = 53
50
51 case 11: 01101010111
52 Likelihood: 0.24384881449471862
53 Beta prior:      a = 78  b = 53
```

■ Case 2: a = 10, b = 1

```
1 case 1: 0101010101001011010101
2 Likelihood: 0.16818809509277344
3 Beta prior: a = 10 b = 1
4 Beta posterior: a = 21 b = 12
5
6 case 2: 0110101
7 Likelihood: 0.29375515303997485
8 Beta prior: a = 21 b = 12
9 Beta posterior: a = 25 b = 15
10
11 case 3: 010110101101
12 Likelihood: 0.2286054241794335
13 Beta prior: a = 25 b = 15
14 Beta posterior: a = 32 b = 20
15
16 case 4: 0101101011101011010
17 Likelihood: 0.18286870706509092
18 Beta prior: a = 32 b = 20
19 Beta posterior: a = 43 b = 28
20
21 case 5: 111101100011110
22 Likelihood: 0.2143070548857833
23 Beta prior: a = 43 b = 28
24 Beta posterior: a = 53 b = 33
25
26 case 6: 101110111000110
27 Likelihood: 0.20659760529408
28 Beta prior: a = 53 b = 33
29 Beta posterior: a = 62 b = 39
30
31 case 7: 1010010111
32 Likelihood: 0.25082265600000003
33 Beta prior: a = 62 b = 39
34 Beta posterior: a = 68 b = 43
35
36 case 8: 11101110110
37 Likelihood: 0.2619678932864457
38 Beta prior: a = 68 b = 43
39 Beta posterior: a = 76 b = 46
40
41 case 9: 01000111101
42 Likelihood: 0.23609128871506807
43 Beta prior: a = 76 b = 46
44 Beta posterior: a = 82 b = 51
45
```

```
46 case 10: 110100111
47 Likelihood: 0.27312909617436365
48 Beta prior:      a = 82  b = 51
49 Beta posterior: a = 88  b = 54
50
51 case 11: 01101010111
52 Likelihood: 0.24384881449471862
53 Beta prior:      a = 88  b = 54
54 Beta posterior: a = 95  b = 58
```