Naive Bayes and Bayesian inference

Aldo N. Cao Romero

May 2022

1 Use case: Phone number scanner App

It is presented the creation of a mobile app used to recognize phone numbers. This app can be useful in many cases, for example, when someone has written a phone number in a piece of paper or even in your hand, when you need to take a quick note of a phone number, or even for a data base where all the phone numbers are written in a list. In addition, this app can also be useful to add bank accounts in a safe way without directly reading and typing the numbers. In this paper we explain how this app works and its efficiency by recognizing numbers.

First, we can observe the main page of the app in figure (1). By typing the "Phone number" key, the app will ask for permission to have access to the contact list, if it is declined, the app won't be able to proceed. In contrary case, the scanner will appear and once the app recognizes the numbers it will give the option of adding to the contact list. The user will choose if he only wants to save the number or adding to his contact list with a name or even a photo. The procedure is analogous to the account number and wi-fi password

Figure 1: Main page of the Phone number scanner app.

The procedure of how the scanner works is the following. We used the training data base showed in figure (2). Then the columns Unnamed:0 and index were dropped. The column labels correspond to the actual values of the numbers which are encoded in the next 784 columns and 60,000 rows.

Once the dropping and the encoded is done, we proceed to obtain an image of the values, some examples can be seen in figure (3)

Then, we apply the Gaussian-Bayes algorithm to the normalized training data. We got an accuracy of 93%. Additionally, and image of the confusion matrix can be seen in figure (4).

	Unnamed: 0	index	labels	0	1	2	3	4	5	6	...	774	775	776	777	778	779	780	781	782	783
0	0	0	5	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	...	0	0	0	0	0	0	0	0	0	0
2	2	2	4	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	3	3	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	4	4	9	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Figure 2: Table of the data.

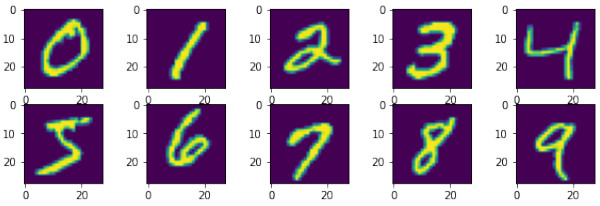

Figure 3: Example of the numbers got from the labels.

Now, the same algorithm was applied to the test normalized Data with 10,000 rows and still with 784 columns, given as a result an accuracy the 99%. The overall of the predicted results can be seen in the confusion matrix with a heat-map in the figure (5).We can see that there were some discrepancies among the number 9. For example, it seems it was principally confused by the numbers 4 and 7. Figure (6).

Nevertheless, we still have a good prediction, in other words, at least in the case of the 9, we are getting 20 bad predictions out of the 10000 rows of data. So, our algorithm is working efficiently.

Finally, we can see the simplicity of the App, giving the opportunity to be useful for many people of different ages. Additionally, due to its simplicity, the uploading of the app ensures the it will be fast, in this manner it will become a practical app. And, at the same time, its development will be of low cost. Also, in order to have profit from it, the option of advertisements is inevitable, however these advertisements wouldn't be pop-up type, in order to avoid being annoying for the user. It is highly recommended to invest in this app since we are dealing with a innovative new technology such as number recognition.

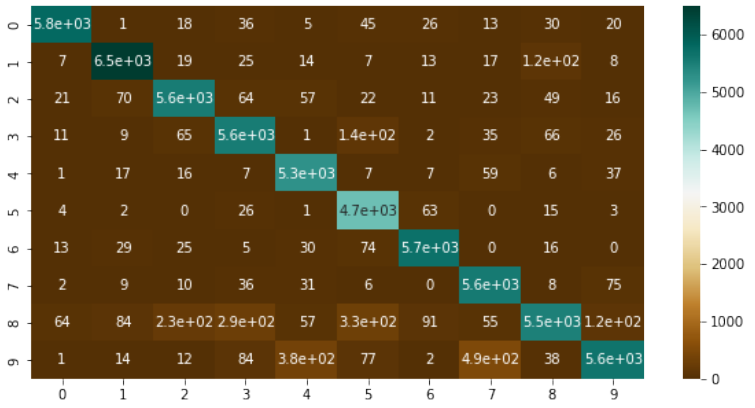

Figure 4: Confusion matrix of the training data.

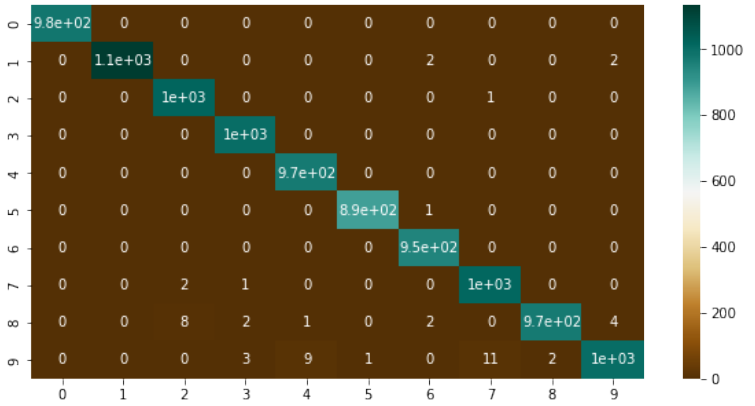

Figure 5: confusion matrix in a heat-map format of the Test data.

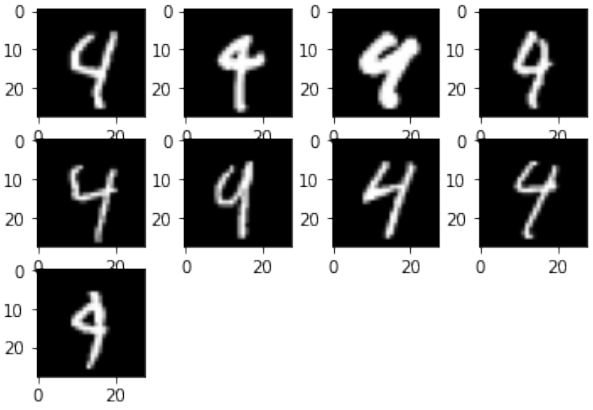

Figure 6: Number 4 confused with the number 9 in the prediction.