Berke Ataseven

DASC521 HW#3

In this homework, we were asked to implement naïve Bayes' classifier. I chose MATLAB as my coding platform. I have converted the given data labels format (A, B, C…) to (1,2,3…) as it is easier to work with integers when an equation is involved. I have estimated my prior class densities and class prior probabilities by implementing the equations given below:

$$\hat{P}(C_i) = \frac{\sum_t r_i^t}{N} \qquad\qquad \hat{p}_{ij} = \frac{\sum_t x_j^t r_i^t}{\sum_t r_i^t}$$

Then I have reshaped my estimated $\hat{p}_{ij}$ vector to 20 by 16 before printing them on a canvas as shown in the figure below:

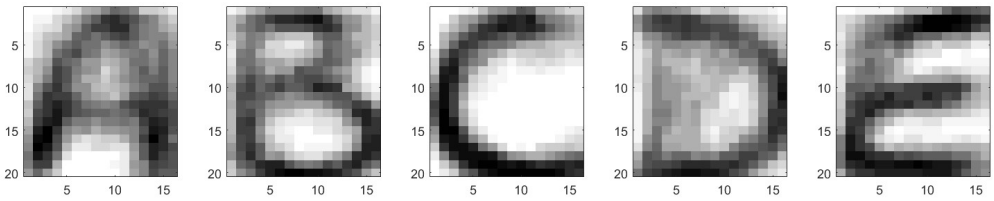

Figure 1 Estimated $\hat{p}_{ij}$'s

Using these values, I have evaluated my discriminant function for each data point – class combination then, to predict the class, I pick the highest-valued class for a single data point. Discriminant function is given below:

$$\begin{aligned} g_i(\boldsymbol{x}) &= \log p(\boldsymbol{x}|C_i) + \log P(C_i) \\ &= \sum_j \left[x_j \log p_{ij} + (1 - x_j) \log(1 - p_{ij})\right] + \log P(C_i) \end{aligned}$$

I calculated score values for both the training test and the testing set. Confusion matrices resulted as shown below:

```
Training_data_ConfusionMatrix =

    25     0     0     0     0
     0    24     0     1     0
     0     1    24     0     0
     0     0     0    25     0
     0     1     0     0    24
```

```
Testing_data_ConfusionMatrix =

    10     1     0     3     0
     0    11     0     3     0
     0     2     8     3     1
     0     2     0    12     0
     0     2     0     0    12
```

I have commented my codes to make it more readable.