Lab exersizes

1

A

Most frequent:

Validation: 14%

Testing: 14%

Naïve bayes:

Validation: 69%

Testing: 55%

Smoothing parameter: 2.00000000

Most frequents correct rate depends entirely on the amount of options and the frequency of said option, 14% could very well be achieved by having 8 options with equal chance, or 2000 options where one pops up 14% of the time. Naïve bayes will do a lot better because it also looks at the properties of said options, it doesn’t relate those to eachother so the degree of error is still quite high. For instance measuring holes in numbers will achieve the same amount of holes, height and with for both 6 and 9 making it prone to errrors for those kind of cases.

B

Laplace smoothing makes sure that if a single property is particularly weird that doesn’t affect the total chance too badly. This makes sure that if something is very out of the ordinary it will still look at the other properties instead of choosing something somewhat at random.

C

This option tries a large array of smoothing parameters to see which one works best and then uses the optimal smoothing parameter. this will generally lead to better results, which it also does here in comparison to a. (74 compared to 69)

D

There be 784. Log 784 is less then 784.

E

Validation is probs on the data that you tuned it on, test on different data, so you see whether it’s overfitted and stuff

5

a very dependent on what the goal is and how much you need to have right, and the amount of time that may be consumed. Generally I would say no because of the low accuracy.

B This command show's which features have the highest odds ration for label 4 over label 2, so the features label 2 is least likely to have and label 4 is most likely to have. Odds are a funny thing we humans use to indicate in how many cases of something happening a certain outcome will present itself. I would interpret the output as the places where if a pixel has a value of higher then 0 the odds of it being a 4 raise substantially.

c ammount of holes, this one is easy, a lot of digits simply have holes or don't so generally it is a good tool to distinguish between digits, unless someone has particularly bad handwriting. The weightpoint of a digit(where you would have to hold your finger below it so it balances), this is rather unique aswell and for a lot of digits isn't in the middle, it also can't be picked up by the features we have so far(this can easily be calculated with physics formula's). the ammount of horizontal blocks(for instance 4×3), this indicates wether a didget has a large straight horizontal part, like a 7, 5 or 4, implementation trivial. Same for vertical block.

d just the holes, this was enough to up percentage to 84%, although it does come with a significant performance decrease. Luckily this wasn't in the exercise so points shouldn't be deducted. It is a, for the extra time it takes, small impact on the end result. Although, again, if accuracy is more important then speed it is still an improvement. The reason for this is as described above that it is a very good tool to distinguish between digits with and without holes.

6

a 88% and 84%