

Assignment -2

Due: Sep 26, 2021 11:59 PM

-
- Please type the solutions using a word processor such as MS Word, Latex, or write by hand neatly and upload the scanned copy of it.
 - I, Andrick (sign your name here), guarantee that this homework is my independent work and I have never copied any part from other resources. Also, I acknowledge and agree with the plagiarism penalty specified in the course syllabus.
 - Turn in your assignment before the deadline. Penalty will be applied to late submission.
-

$$1. p(c = \text{Black} | \text{good})$$

$$= \frac{6}{17} = 0.35$$

$$p(c = \text{green} | \text{good})$$

$$= \frac{6}{17} = 0.35$$

$$p(c = \text{white} | \text{good})$$

$$= \frac{5}{17} = 0.29$$

2. age < 30 , income = medium, student = yes, credit = fair

	< 30	Medium	Fair	total
NO	3	2	2	5
Yes	2	4	6	9
total	5	6	8	14

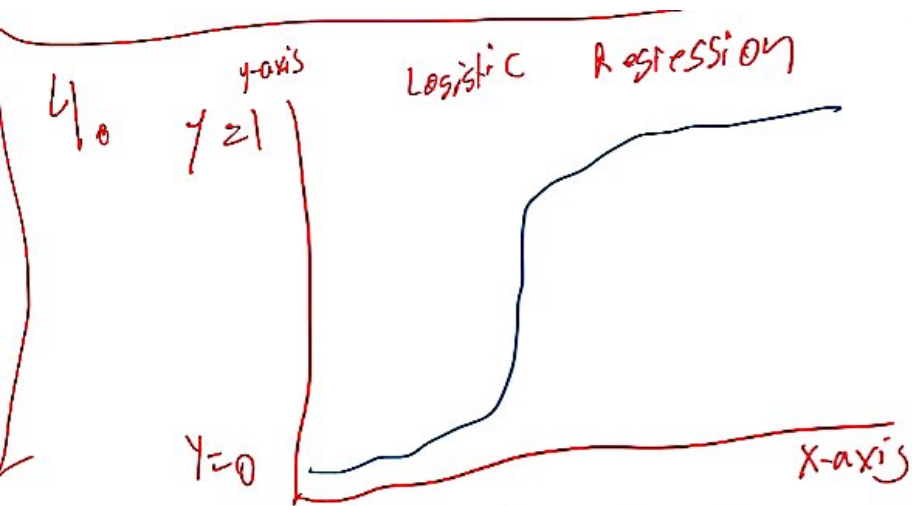
$$P(\text{NO} | c=30, \text{medium}, \text{yes}, \text{fair}) = \left(\frac{5}{14}\right) \cdot \left(\frac{3}{5}\right) \left(\frac{2}{5}\right) \left(\frac{1}{5}\right) \left(\frac{2}{5}\right) = 0.0068$$

$$P(\text{Yes} | c=30, \text{medium}, \text{yes}, \text{fair}) = \left(\frac{9}{14}\right) \cdot \left(\frac{2}{9}\right) \left(\frac{4}{9}\right) \left(\frac{6}{9}\right) \left(\frac{6}{9}\right) = 0.628$$

∴ This particular person will buy a

computer

3. To prevent this from
occurring we can add
any constant value such
as 1 to all training
data, to avoid 0's

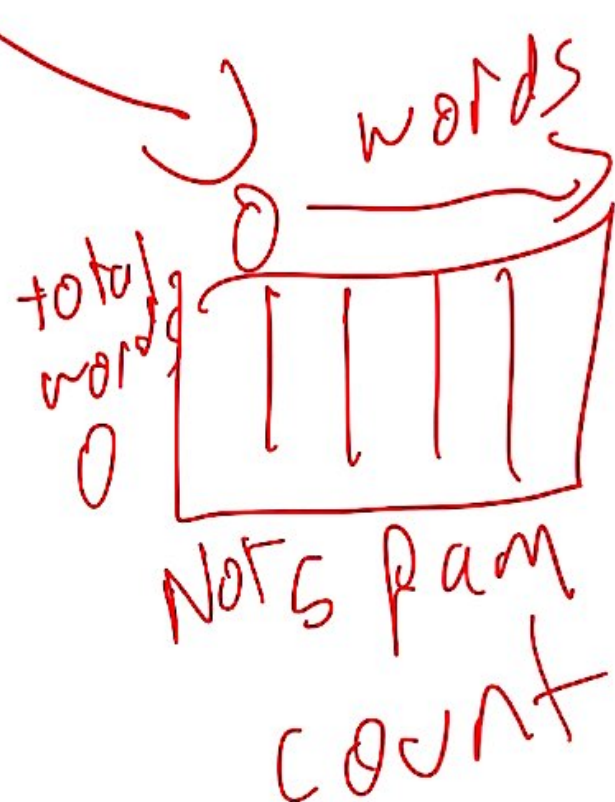
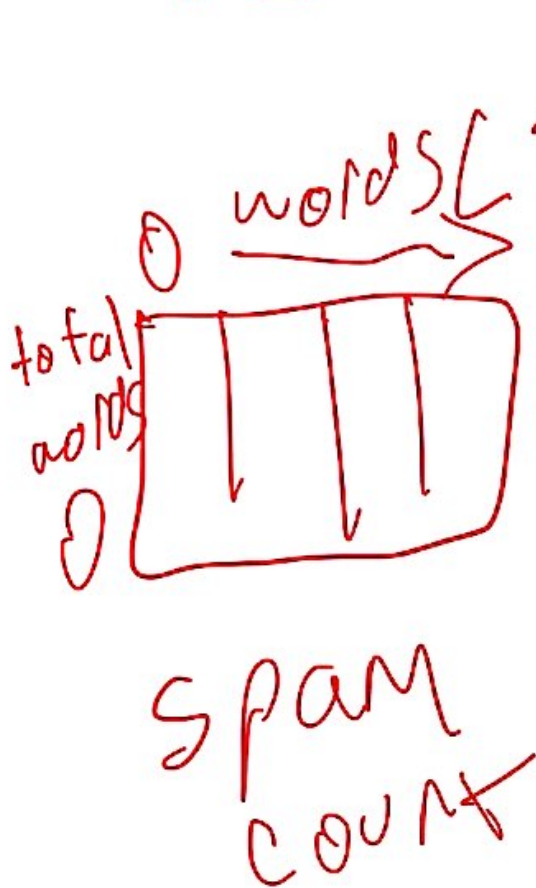
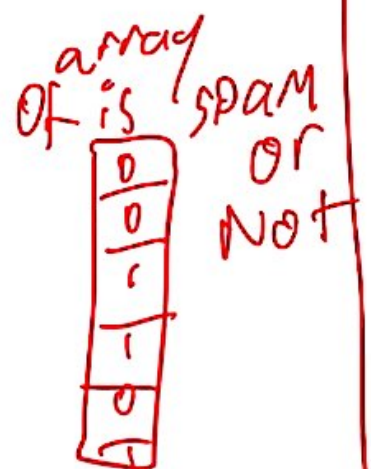
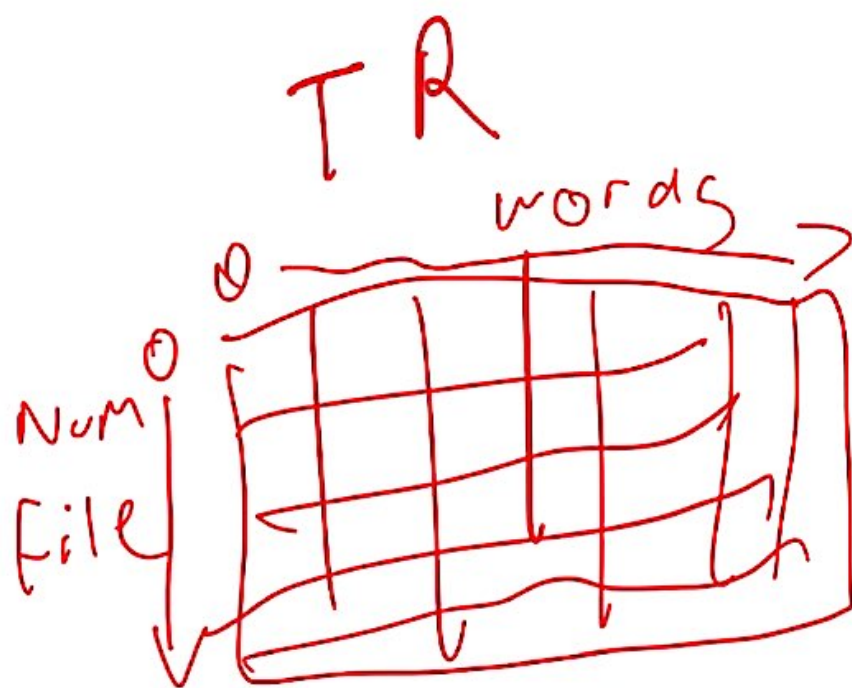


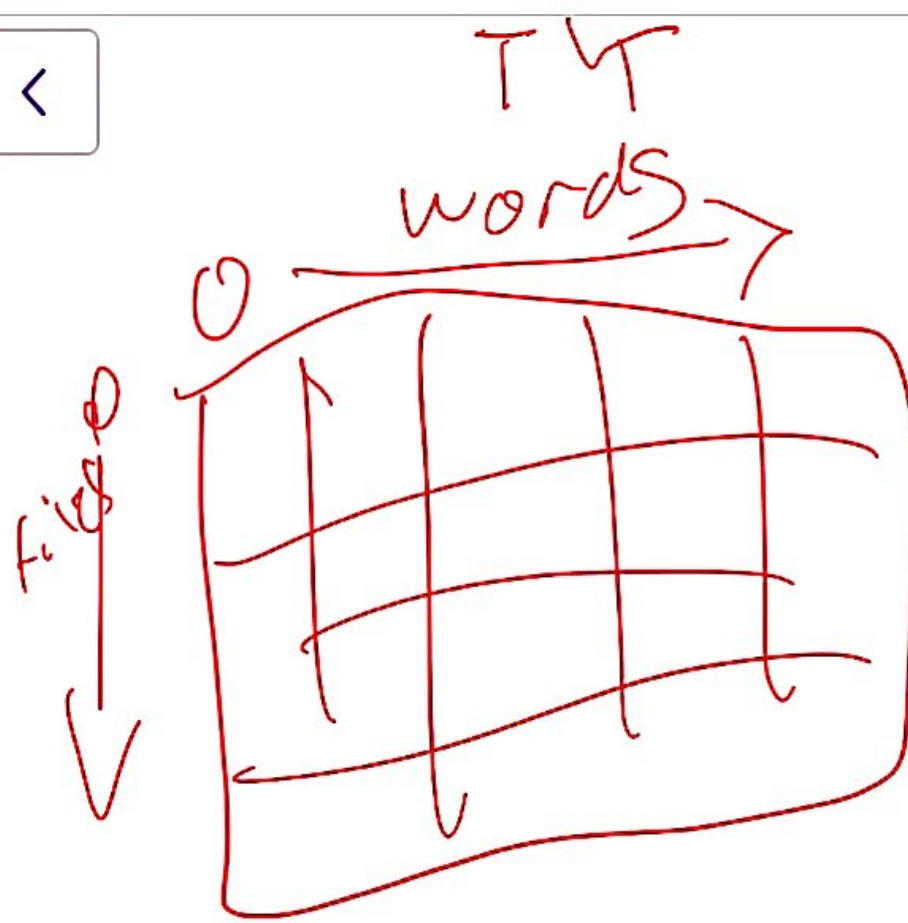
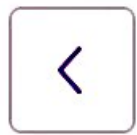
- The sigmoid function allows to model predicted values to probabilities, essentially turns any value to either 0 or 1,
- Cost and gradient descent allow us to put cost based on error of our calculations, by correlating the two we can optimize the algorithm

Part B program

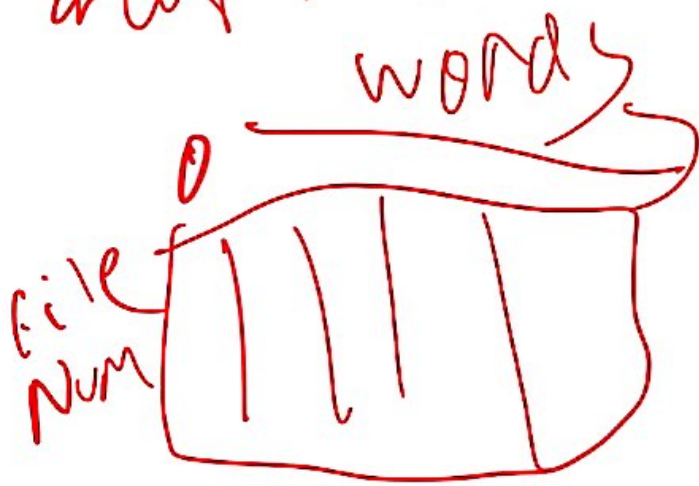
- I split the data by checking the sum of all columns that were spam or 1, and then the rest of the data was sent to not spam or 0
- once data was split I checked if the word was contained in the training data, did the calculations in an array

my solution is not well optimized but it still works





after one file iteration



from here we predict
if its either spam or
not spam

< 1000

• we predict by
appending probabilities
to an array, and
then calculating
the product of the array
if the probability of
spam is greater, than
the file is spam
vice versa
