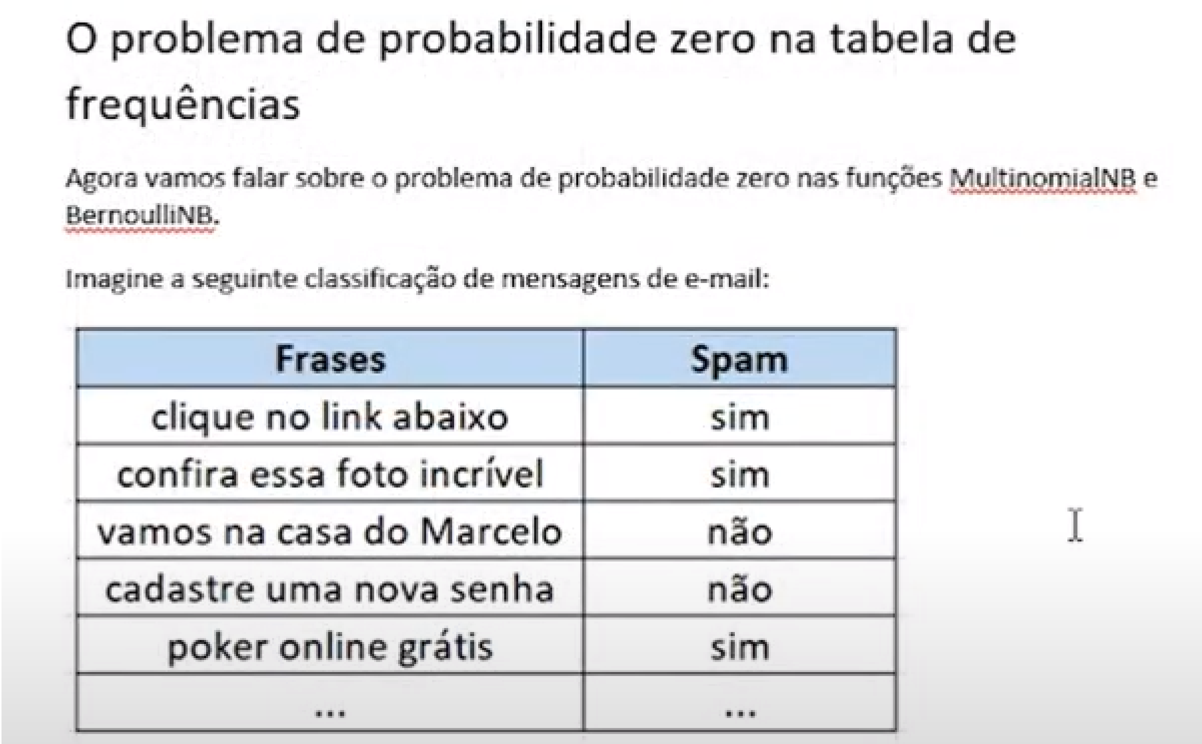
Suavização de Laplace

Basicamente ela serve para os problemas que utilizam MultinomialNB e BernoulliNB, onde os valores utilizados são discretos. Podemos ver uma aplicação disso no exemplo abaixo:



O problema de probabilidade zero na tabela de frequências

Agora vamos falar sobre o problema de probabilidade zero nas funções MultinomialNB e BernoulliNB.

Imagine a seguinte classificação de mensagens de e-mail:

Frases	Spam
clique no link abaixo	sim
confira essa foto incrível	sim
vamos na casa do Marcelo	não
cadastre uma nova senha	não
poker online grátis	sim
...	...

Podemos ter um dataset com várias mensagens classificadas com Spam e Não Spam, e a partir disso utilizar o teorema de Bayes para encontrar a probabilidade de uma nova mensagem ser considerada Spam ou não. Para isso os eventos utilizados na fórmula, serão as palavras em sí. Podemos contabilizar quantas vezes uma palavra aparece em uma mensagem de spam como abaixo:

Tabela de Frequências		
Palavra	*Spam = sim*	*Spam = não*
clique	68	3
no	42	115
link	13	5
abaixo	7	1
confira	9	0
vamos	2	76
...	...	...

P(sim/"confira essa foto louca") = P(sim/(confira, essa, foto, louca))

Agora exemplificando com a probabilidade da frase "confira essa foto louca" ser Spam ou não Spam, podemos aplicar na fórmula:

P(sim/"confira essa foto louca") = P(sim/(confira, essa, foto, louca))

P(A/B) = P(B/A)*P(A) / P(B)

P(A/(x1, x2, x3, x4)) = P(x1/A)*P(x2/A)*P(x3/A)* P(x4/A)*P(A) / P(x1)*P(x2)*P(x3) *P(x4)

Repare que quando uma palavra não existe na tabela de frequências, isso resultaria em probabilidade zero.

Como podemos ver a palavra louca não existe na tabela de frequência, e isso causaria um problema nas multiplicações, pois a probabilidade seria 0, o que cancelaria todas as outras probabilidades.

Para resolver esse problema de amostras que não continha nos dados previamente identificados, podemos utilizar a suavização de Laplace.

O método utilizado para contornar essa situação é a **suavização de Laplace.** Essa suavização é dada pela fórmula:

$$\hat{\theta}_i = \frac{x_i + \alpha}{N + \alpha d} \qquad (i = 1, \ldots, d),$$

Onde Theta-i é o novo parâmetro de cálculo de probabilidade (suavizado), xi são as observações desse parâmetro, alfa é o suavizador (default = 1), N é o total de ocorrências (tabela de frequência) dos parâmetros e d é o total de parâmetros.

Abaixo há um exemplo de aplicação da suavização:

Utilizando como base a tabela de frequências do primeiro exercício:

Tabela de Frequências		
Resultado do jogo	*Feliz = sim*	*Feliz = não*
vitória	6	1
empate	3	2
derrota	1	4
Total	**10**	**7**

Cada probabilidade precisaria ser ajustada. Por exemplo, a probabilidade de vitória, em vez de ser 7/17 = 0,41, ficaria (supondo alfa = 1):

P(vitória) = θ_1 = (7+1)/(17+1*3) = 0,40.

P(empate) = θ_2 = (5+1)/(17+1*3) = 0,30.

P(derrota) = θ_3 = (5+1)/(17+1*3) = 0,30.

Essa fórmula mostra que, quando um dado novo i que nunca apareceu antes precisa ser testado no modelo, em vez de receber probabilidade zero, acaba recebendo a probabilidade de:

$\theta_i = 1/(N + d)$

Quando alfa = 0, o cálculo elimina o fator suavização:

$\theta_i = x_i/N$

Basicamente aplicando a suavização, dado uma amostra que nunca apareceu antes, sua probabilidade será dada por: 0i = 1 / (N + d), e não seria 0.

Quando deixamos o Alfa da suavização zerado, estaremos aplicando a mesma fórmula de probabilidade usada anteriormente, xi / N, onde é o número de vezes que o evento ocorreu dividido pelo número total de eventos.