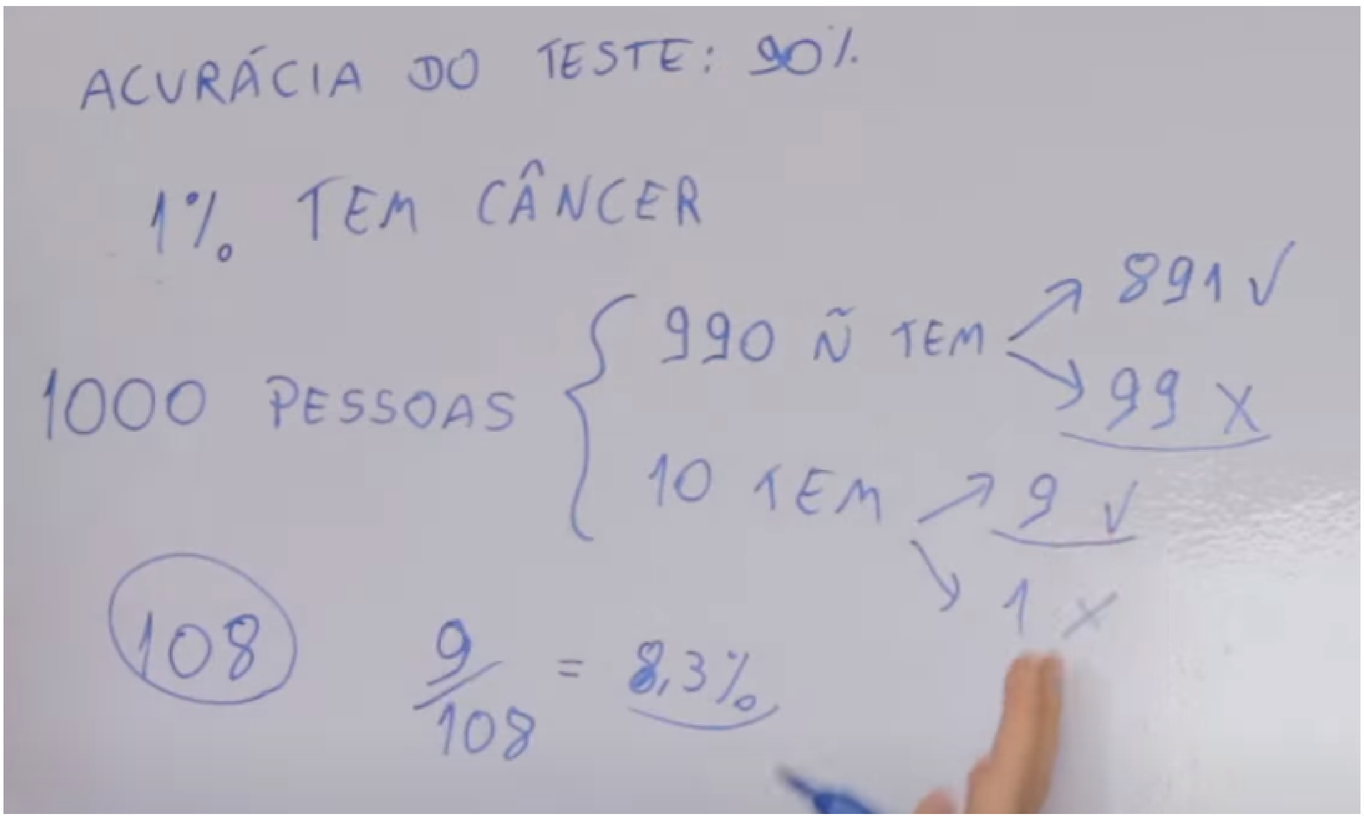
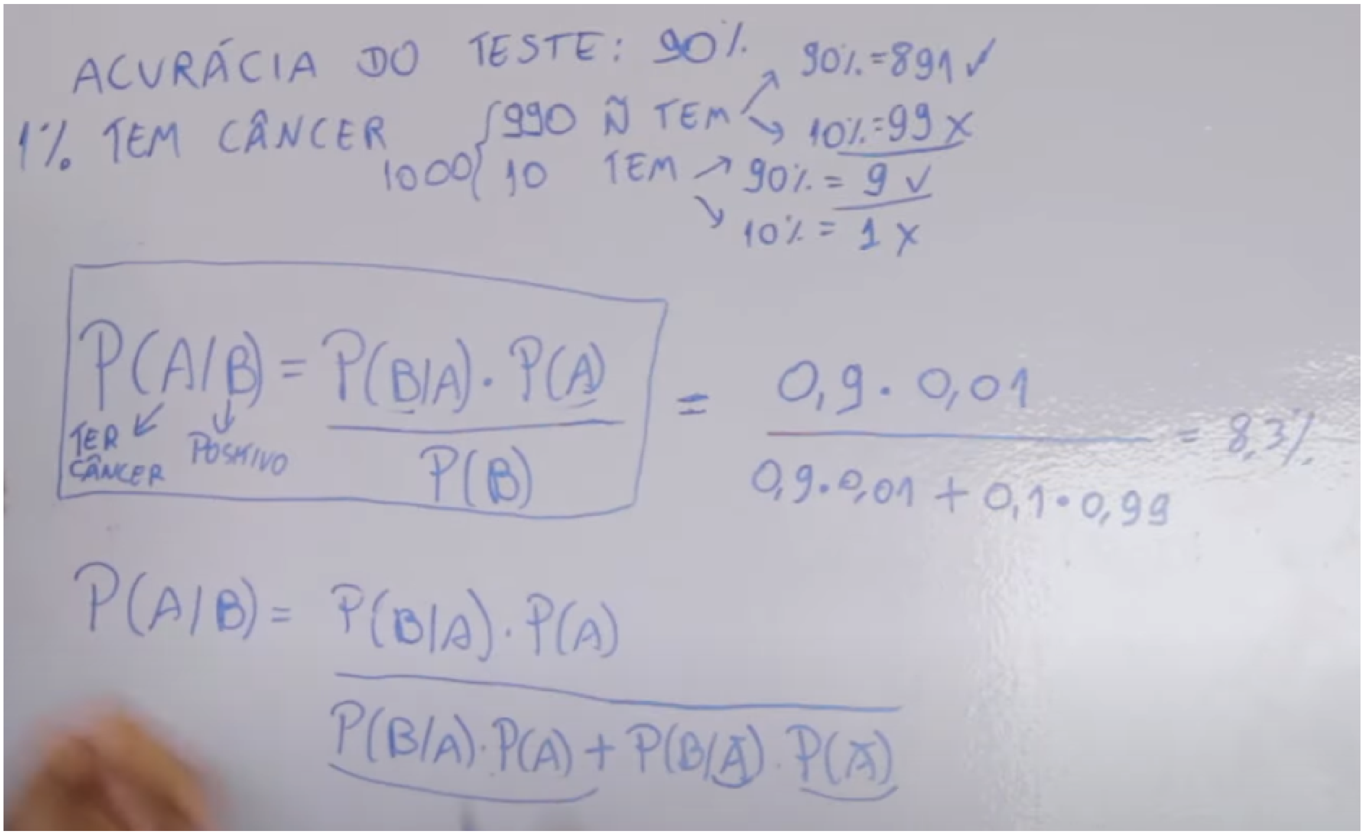
# Teorema de Bayes

Dado um modelo que acerta 90% dos casos de pessoas que não tem o câncer e 90% dos casos que têm o câncer, temos como intuição dizer que a chance de ter câncer é de 90%, porém isso não é o correto, e o teorema de Bayes mostra a maneira correta de se interpretar o problema. Antes de entrar no teorema, abaixo podemos relacionar esse exemplo, onde precisamos primeiro saber qual é o tamanho da população, nesse caso 1000 pessoas, onde 990 não têm o câncer e 10 têm. Usando o modelo, podemos prever que 891 pessoas não tem câncer, e como ele erra em 10%, 99 ele classificaria como um falso positivo. Já nas que têm câncer, acertaria 9 e erraria 1 pessoa. Ao somar 99 + 9 que são as pessoas que foram classificadas com câncer, temos 108 no total, ou seja, a probabilidade do modelo te classificar com câncer é de 8,3% e não 90%.




Agora utilizando o teorema de Bayes, podemos simplificar esse exemplo utilizando a fórmula abaixo:




Devemos seguir a fórmula de probabilidades, basicamente o que queremos descobrir é qual a probabilidade do resultado dar positivo e a pessoa realmente ter câncer. Para isso A é o evento ter o câncer, e B é o evento ser positivo realmente.

A parte de cima da equação nós já temos, sabemos que P(B/A), ou seja, Probabilidade de ocorrer B, dado que A já aconteceu, nesse caso a probabilidade de dar positivo se a pessoa realmente tem o câncer. E isso já está dado na acurácia, se o modelo acerta 90% das vezes, então P(B/A) = 0.9 e P(A) = 0.01, pois somente 1% da população tem câncer.

Agora para a parte debaixo da divisão, no caso P(B) (dar positivo), precisamos destrinchar um pouco as probabilidades de B acontecer. Que seriam, a probabilidade de B acontecer se A já aconteceu *P(A) (Probabilidade de ter câncer) + o oposto, B acontecer caso A ainda não aconteceu* P(Ã) (Não ter câncer). Com isso temos a probabilidade de ser positivo P(B). Resumindo, a probabilidade de B é calculada quando A ocorre e quando A não ocorre.

Portanto no denominador temos P(B/A) *P(A) + P(B/Ã)* P(Ã), 0.9 *0.01 + 0.1* 0.99

0.9, pois há 90% de acerto quando a pessoa têm câncer
0.01, pois somente 1% da população tem câncer
0.1, pois há 10% de chance de erro quando a pessoa não tem câncer e o modelo dizer que têm
0.99, pois de chance da pessoa não ter câncer

Calculando tudo, temos 8,3% de chance de ter câncer se o evento for positivo. O teorema acaba deixando de forma mais genérica com porcentagens da população, e não com a quantidade exata dos dados, ou seja, poderiam ser milhões de registros, mas somente trabalha-se com a porcentagem.

## Quando há mais de uma feature

Nesse caso só temos um evento, mas pensando em machine learning, podemos ter N features em nosso dataset, e devemos considerar que todas elas são eventos que acontecem simultãneamente, o que resulta na fóruma abaixo:

Quando existe mais de uma feature:

$$P(C_k|X) = \prod_{i=1}^{n} P(x_i|C_k)$$

$$P(C_i|x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n|C_i).\,P(C_i)}{P(x_1, x_2, \ldots, x_n)} \text{ for } 1 \leq i \leq k$$

Explicando melhor:

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

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

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